

Emergence of scale-free close-knit friendship structure in online social networks

Ai-Xiang Cui¹, Zi-Ke Zhang^{2,1}, Ming Tang^{1,*}, Pak Ming Hui³, Yan Fu¹

1 Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, People's Republic of China

2 Institute for Information Economy, Hangzhou Normal University, Hangzhou 310036, People's Republic of China

3 Department of Physics, The Chinese University of Hong Kong, Shatin, Hong Kong, People's Republic of China

* E-mail: tangminghuang521@hotmail.com

Abstract

Despite the structural properties of online social networks have attracted much attention, the properties of the close-knit friendship structures remain an important question. Here, we mainly focus on how these mesoscale structures are affected by the local and global structural properties. Analyzing the data of four large-scale online social networks reveals several common structural properties. It is found that not only the local structures given by the indegree, outdegree, and reciprocal degree distributions follow a similar scaling behavior, the mesoscale structures represented by the distributions of close-knit friendship structures also exhibit a similar scaling law. The degree correlation is very weak over a wide range of the degrees. We propose a simple directed network model that captures the observed properties. The model incorporates two mechanisms: reciprocation and preferential attachment. Through rate equation analysis of our model, the local-scale and mesoscale structural properties are derived. In the local-scale, the same scaling behavior of indegree and outdegree distributions stems from indegree and outdegree of nodes both growing as the same function of the introduction time, and the reciprocal degree distribution also shows the same power-law due to the linear relationship between the reciprocal degree and in/outdegree of nodes. In the mesoscale, the distributions of four closed triples representing close-knit friendship structures are found to exhibit identical power-laws, a behavior attributed to the negligible degree correlations. Intriguingly, all the power-law exponents of the distributions in the local-scale and mesoscale depend only on one global parameter – the mean in/outdegree, while both the mean in/outdegree and the reciprocity together determine the ratio of the reciprocal degree of a node to its in/outdegree. Structural properties of numerical simulated networks are analyzed and compared with each of the four real networks. This work helps understand the interplay between structures on different scales in online social networks.

Introduction

In recent years, an increasing number of online social systems (*e.g.*, *YouTube* and *Facebook*) have been attracting wide attention from different fields [1–3]. Online social networks provide a platform for web surfers to make acquaintance with congenial friends [4], exchange photos and personal news [5], share videos [6], establish communities or forums on focused issues [7], etc. These online interactive behaviors, which partly reflect real-life social relationships among people, provide an unprecedented opportunity to study and understand the dazzling characteristics of real-life social systems [8,9].

Complex network theory has been proven to be a powerful framework to understand the structure and dynamics of complex systems [10–16]. Online social systems have been treated as undirected networks [17, 18], which have been applied successfully in exploring various systems [10]. This simplification, however cannot describe the asymmetric interactions among users. Taking *Flickr* as an example, if a user *A* designates another user *B* as a friend, user *A* can see the photos of user *B*, but not the other way round

unless user B also designates user A as his friend. Technically, an asymmetric interaction represents one directed link, and many online social systems are thus directed networks in nature. The directionality of links is important in characterizing the functioning of many systems, *e.g.*, leadership structure of social reputation [19, 20], reciprocal behavior in evolutionary games [21], information hierarchy of the World Wide Web [22, 23], citation relationship of scientific publications [24, 25], etc. Much effort has been devoted to understanding the structural properties of these directed networks, including the indegree and outdegree distributions [26], average shortest distance [26], degree correlation [27], and community structure [28–30]. Correspondingly, there are many models proposed for the underlying mechanisms of the statistical properties. Dorogovtsev *et al.* [31] generalized the Barabási-Albert (BA) model [32] and obtained the exact form of the indegree distribution of growing networks in the thermodynamic limit. Krapivsky *et al.* [33] introduced a directed network model that generates correlated indegree and outdegree distributions. Zhou *et al.* [20] argued that the “good get richer” mechanism would facilitate the emergence of scale-free leadership structure in online social networks.

Up to now, most of the work on complex networks can be classified into studies on three scales: the local scale based on the single node properties (through statistical distributions), the macro-scale based on the global properties of networks (with global parameters), and the mesoscale based on properties due to a group of nodes (via modular properties) [34–36]. However, a majority of studies focused on the first two scales. In view of the significant role of modularity in the functionality of real networks, it has become increasingly important to study the mesoscale structures. Communities and motifs are two key mesoscale structures of real complex networks. Community structures at mesoscale level are ubiquitous in a variety of real complex systems [37, 38], such as *Facebook*, *YouTube*, and *Xiaonei*. There are more connections among members of the same community than among members in different communities. Lancichinetti *et al.* analyzed the statistical properties of communities in five categories of real complex networks, and found that communities detected in networks of the same category display similar structural characteristics [39]. Motifs, which are defined as subgraphs that occur much more often than expected in a random network, play a significant role in our understanding of the interplay between the structures and dynamics of real complex networks [40–45].

In spite of the structural features revealed at the three scales, understanding the interplay between the different scales has remained a major challenge [34–36]. In the present work, we study how the close-knit friendship structures of online social networks at the mesoscale level and the structural properties at the two other scales are affecting each other. In social networks, the close-knit friendship structure describes the closest unit, which is usually represented by the closed triples. In a directed network, there are 13 different possible three-node subgraphs [41]. For situations without reciprocal links, a focal node has three possible unclosed triples. Each unclosed triple can be closed by adding a directed link between the two unconnected nodes, giving rise to four types of closed triples as shown in Figure 1 [44, 45]. The four closed triples fall into two groups: one is a feedback (FB) loop and the three others are feedforward (*i.e.*, FF_a , FF_b , and FF_c) loops. Structurally, the roles of three nodes in the FB loop are equivalent, but it is not the case in the FF loops. Any FF_a loop (from the perspective of the focal node) becomes a FF_b loop for another node and a FF_c loop for the third node, and thus the numbers of three feedforward loops are equal in directed networks. Compared to the unclosed triples, the closed triples play a more important role in dynamical processes on online social networks [46, 47], such as opinion formation [48], game dynamics [49], and cooperation evolution [50].

In online social networks, the closed triples are a good indicator of close-knit friendships among people. To understand the mesoscale structural properties of online social networks, we analyze data of popular online social networks, establish the empirical facts, and introduce a directed network model. We analyze four large-scale online social networks, namely *Epinions*, *Slashdot*, *Flickr*, and *Youtube*, and establish that the distributions in each scale follow a similar power law. We propose a simple directed network model incorporating two processes: external reciprocation and internal evolution. Theoretical analysis shows that the distributions of four closed triples display almost identical scaling laws due to the negligible

degree correlations, and the distribution exponents depend only on one global parameter - the mean in/outdegree. Simulation results based on the model are basically consistent with both the empirical results and theoretical analysis.

Results

Empirical Results

We first analyze four representative directed online social networks and establish the empirical features. As listed in Table 1, these four datasets are: (i) *Epinions* Social Network (ESN, <http://snap.stanford.edu/data/soc-Epinion>) a who-trust-whom online social network of a general consumer review site *Epinions.com* in which members can decide whether to “trust” each other or not, and subsequently all the trusted relationships form a so-called social trust network. (ii) *Slashdot* Social Network (SSN, <http://snap.stanford.edu/data/soc-Slashdot0902.html>) [51]: a friendship network of a technology-related news website *Slashdot.com*. Nodes are the users and links represent the friendships among the users. (iii) *Flickr* Social Network (FSN, <http://socialnetworks.mpi-sws.org/data-imc2007>) a friendship network of a photo-sharing site *Flickr.com* that allows users to designate others as “contacts” or “friends” and track their activities in real time. This network contains all the friendship links among the users of Flickr. (iv) *YouTube* Social Network (YSN, <http://socialnetworks.mpi-sws.org/data-imc2007.html>) [52]: a friendship network of a popular video-sharing website *YouTube.com* on which users can upload, share and view videos. The nodes in the network are the users of YouTube, and a directed link is established from a user A to a user B when user A declares user B as a friend. Table 1 summarizes the basic global features of the four online social networks. These networks all show a large reciprocity r , defined by $r = E_r / (E - E_r)$ [53] with E_r and E being the numbers of reciprocal links and single directed links, respectively. Note that a reciprocal link contributes two single directed links. For example, $r \approx 0.25$ for ESN, $r \approx 0.73$ for SSN, $r \approx 0.45$ for FSN, and $r \approx 0.65$ for YSN.

We also studied the local-scale structural properties of these social networks via statistical distributions. The results of ESN are presented as an example. Figure 2 shows the indegree and outdegree distributions (black squares) on a log-log plot. The data span more than two decades. The distributions follow a power law with approximately the same exponent, *i.e.*, $P(k_{in}) \sim k_{in}^{-\gamma_{in}}$ and $P(k_{out}) \sim k_{out}^{-\gamma_{out}}$, with $\gamma_{in} \approx 1.73$ and $\gamma_{out} \approx 1.71$ obtained by the maximum likelihood estimation [54, 55]. More details about the power-law fits are given in Table S1 of Appendix SI. Figure 3 shows that the indegree k_{in} of each node is nearly proportional to its outdegree k_{out} (also see Figures S4-S6 of Appendix SI), which is consistent with the similar scaling law of their distributions. In growing networks, the fat-tail power-law behavior in the degree distribution suggests that directed links are not drawn toward and from existing users uniformly. Mislove *et al.* showed that there is a positive correlation between the number of links a user has and its probability of creating or receiving new links in online social networks [5]. This phenomenon is called “preferential attachment” [5, 32, 33, 56]. The behavior $k_{in} \approx k_{out}$ for any node implies that a node with large k_{in} has a strong ability to attract links from other nodes and also a strong tendency to link to other nodes. This is reminiscent of the product $k_{out}^i k_{in}^j$ used in the prediction of a link between the nodes i and j [57], *i.e.*, a larger product gives a larger probability of having a directed link from i to j . These results lead us to incorporate a preferential attachment mechanism related to $k_{out}^i k_{in}^j$ into the mechanism of how the links grow in a network.

The reciprocal degree is the number of reciprocal links that a node possesses. Figure 4 shows that the reciprocal degree distribution also follows a power law $P(k_r) \sim k_r^{-\gamma_r}$ with an exponent $\gamma_r \approx 1.69$ as examined by the maximum likelihood estimation [54, 55], similar to that of the indegree and outdegree distributions. Figure 5 shows that the mean reciprocal degree of the nodes with the same indegree $\langle k_r(k_{in}) \rangle$ is approximately linearly proportional to the indegree k_{in} (also see Figures S10-S12 of Appendix SI), *i.e.*, $\langle k_r(k_{in}) \rangle \sim k_{in}$, and in a similar fashion $\langle k_r(k_{out}) \rangle \sim k_{out}$, implying that the probability that a randomly chosen directed link happens to be a reciprocal link is roughly a constant. All these features

are consistent with the observation that the indegree, outdegree, and reciprocal degree distributions all follow a similar exponent.

For mesoscale structures, we focus on the four closed triples *i.e.*, FB , FF_a , FF_b and FF_c . As the numbers of three feedforward loops are equal, *i.e.*, $N_{FF_a} = N_{FF_b} = N_{FF_c}$, we only look at the total numbers of FB and FF_a closed triples. For ESN, $N_{FB} = 740,310$ and $N_{FF_a} = 3,586,403$ as shown in Table 1. Considering the feedforward loops as the same up to the permutation of the focal node, it is interesting to see that $N_{FB} : N_{FF_a} \approx 1 : 5$. This implies the existence of some underlying mechanism. Since the indegree and outdegree distributions are heterogeneous, we study the numbers of the four closed triples (*i.e.*, n_{FB} , n_{FF_a} , n_{FF_b} , and n_{FF_c}) at different nodes and their distributions. Figure 6 shows that, although the numbers of feedback and feedforward loops are different, their distributions follow similar scaling laws, *i.e.*, $P(n_{FB}) \sim n_{FB}^{-\gamma_{FB}}$ and $P(n_{FF}) \sim n_{FF}^{-\gamma_{FF}}$, with $\gamma_{FB} \approx 1.37$, $\gamma_{FF_a} \approx 1.39$, $\gamma_{FF_b} \approx 1.35$ and $\gamma_{FF_c} \approx 1.38$ as determined by the maximum likelihood estimation [54, 55]. More details on the exponents are given in Table S1 of Appendix SI. Moreover, although the numbers of three feedforward loops are equal, their distributions look slightly different in detail. This is a phenomenon worthy of further research.

To understand this phenomenon, we consider the three unclosed triples in Figure 1. For a node with indegree k_{in} and outdegree k_{out} , there are $C_{k_{out}}^2$ unclosed triples A , $C_{k_{in}}^1 C_{k_{out}}^1$ unclosed triples B , and $C_{k_{in}}^2$ unclosed triples C when reciprocal links are forbidden, where $C_n^m = n!/[m!(n-m)!]$ denotes the binomial coefficient. These unclosed triples would generate closed triples in the ratio $n'_{FB} : n'_{FF_a} = (C_{k_{in}}^1 C_{k_{out}}^1/2) : (C_{k_{out}}^2 + C_{k_{in}}^2 + C_{k_{in}}^1 C_{k_{out}}^1/2)$. Accounting for all the nodes, we can obtain the total number of optional closed triples $N'_{FB} = \sum_i (C_{k_{in}}^1 C_{k_{out}}^1/2)$ and $N'_{FF_a} = \sum_{i=1}^N (C_{k_{out}}^2 + C_{k_{in}}^2 + C_{k_{in}}^1 C_{k_{out}}^1/2)$, respectively. Assuming there is no degree correlation and making use of $k_{in} \approx k_{out}$, we have $N_{FB} : N_{FF_a} \approx 1 : 5$, which is basically consistent with the ratio found in ESN. The assumption of no degree distribution is supported by the results in Figure 7(a), in which the network shows a very weak degree correlation over two decades that can be treated almost as no degree correlation (further quantitative evidence is given by the Pearson correlation coefficient in Table S2 of Appendix SI) [58]. In this case, the number of closed triples at a node depends only on its indegree k_{in} and outdegree k_{out} , *i.e.*, $n_{FB} \sim k_{in}^2$ and $n_{FF_a} \sim k_{in}^2$ for large k_{in} , $n_{FB} \sim k_{out}^2$ and $n_{FF_a} \sim k_{out}^2$ for large k_{out} . This behavior is confirmed in Figure 8 and Figure 9 (also see Figures S19-S24 of Appendix SI). This also gives the reason why the distributions of four closed triples follow similar scaling laws. Results of analyzing the other three networks (*i.e.*, *Slashdot*, *Flickr* and *YouTube*) also exhibit similar phenomena (see Figures S1-S24 of Appendix SI).

Directed Network Model

We propose a growing network model with node and link creation processes incorporating link directionality that reproduces the empirical features. In the model, we consider two evolutionary ingredients: reciprocation and preferential attachment. On one hand, many empirical results show that the reciprocity r of online social networks is much greater than in sparse random directed networks with $r \rightarrow 0$ [5, 53]. Our results of $r \approx 0.45$ of FSN and $r \approx 0.65$ of YSN provide further evidence. The high reciprocity implies that there is a good chance that the creation of a directed link prompts the establishment of a reversed link. For example, users of Flickr often respond to an incoming link by quickly establishing a reversed link as a matter of courtesy [5]. Thus, reciprocation is believed to be an independent growth mechanism in large-scale online social networks. On the other hand, preferential attachment has been proven to be an important and basic growing mechanism in online social networks [5, 32, 33, 56]. Users with large indegrees and outdegrees are more likely to receive incoming links and create outgoing links, respectively. This motivated us to incorporate a preferential attachment mechanism depending on the product $k_{out}^i k_{in}^j$ in creating new links.

The model starts with an initial seed consisting of m_0 nodes. At each time step, a new node is added and $2 + m + mp$ new directed links are introduced according to two processes: external reciprocation and

internal evolution.

(1) External reciprocation. The new node in every time step establishes a new directed link with an existing nodes i in the network with a probability

$$p_i = \frac{k_{in}^i}{\sum_j k_{in}^j} \quad (1)$$

proportional to the indegree k_{in}^i of node i . To incorporate the reciprocation mechanism, the node i that receives the link creates a reversed link to the new node. Consequently, a reciprocal link is created between these two nodes. This mechanism is reasonable in that a strong motivation of a new user joining a social network is to get connected to and interact with someone already in the network. As we shall see, this process can be treated conveniently in the mathematical analysis of the model.

(2) Internal evolution. In each time step, m new directed links, representing the activity of the network, are created among the existing nodes according to the preferential attachment mechanism. Consider two unconnected nodes i and j up to that time step, a new directed link from node i to node j is created with the probability

$$p_{ij} = \frac{k_{out}^i k_{in}^j}{\sum_{x,y, x \notin \Gamma_{in}(y)} k_{out}^x k_{in}^y}, \quad (2)$$

where k_{out}^i and k_{in}^j are the outdegree of node i and the indegree of the target node j , respectively, and $\Gamma_{in}(y)$ in the normalization factor is the set of incoming neighbors of node y at that time step. This attachment probability is proportional to the product $k_{out}^i k_{in}^j$. The larger the product is, the greater probability a new directed link is created between them. For each of the new directed links created, a reversed link will be established with the reciprocation probability p . Therefore, $m + mp$ directed links are introduced into the network through internal evolution in each time step. It should be noted that multiple links between two nodes and self-connections are prohibited in the model.

Materials and Methods

Rate Equation Analysis

We first analyze the indegree and outdegree distributions of the model. After t steps, the growing directed network has $N = m_0 + t$ nodes and $(2 + m + mp)t$ directed links, where the tiny number of initial links in the seed are ignored. Meanwhile, the sum of indegree and the sum of outdegree are equal, *i.e.*, $\sum_j k_{in}^j = \sum_j k_{out}^j = (2 + m + mp)t$. For a sparse network with mean indegree $\langle k \rangle = 2 + m + mp \ll N$, we have $\sum_{x,y, x \notin \Gamma_{in}(y)} k_{out}^x k_{in}^y \approx \sum k_{out}^x \times \sum k_{in}^y = [(2 + m + mp)t]^2$ so that Eq. (2) can be approximated by

$$p_{ij} \approx \frac{k_{out}^i k_{in}^j}{[(2 + m + mp)t]^2}. \quad (3)$$

Consider the creation of one new directed link via the internal evolution at step t . The probability $p_{k_{in}^i}^+$ that the indegree k_{in}^i of node i increases by one due to the creation of one link is

$$p_{k_{in}^i}^+ = \sum_{j \notin \Gamma_{in}(i)} \frac{k_{out}^j k_{in}^i}{[(2 + m + mp)t]^2} + p \sum_{j \notin \Gamma_{out}(i)} \frac{k_{out}^i k_{in}^j}{[(2 + m + mp)t]^2}, \quad (4)$$

where the first term gives the probability that the node i receives a new incoming link from one of the other nodes and the second term gives the probability that a reversed link is created back to node i when a

new directed link was created from node i to some node j . According to $\sum_{j \notin \Gamma_{in}(i)} k_{out}^j \simeq \sum_{j \notin \Gamma_{out}(i)} k_{in}^j \simeq (2 + m + mp)t$, $p_{k_{in}^i}^+$ is approximately given by

$$p_{k_{in}^i}^+ \approx \frac{k_{in}^i + pk_{out}^i}{(2 + m + mp)t}. \quad (5)$$

Similarly, the probability $p_{k_{out}^i}^+$ that the outdegree k_{out}^i of node i increases by one due to the creation of one link is

$$p_{k_{out}^i}^+ \approx \frac{k_{out}^i + pk_{in}^i}{(2 + m + mp)t}. \quad (6)$$

Equations for the rate of change of the expected indegree k_{in}^i and outdegree k_{out}^i can then be written down. Taking k_{in}^i and k_{out}^i as continuous variables, the dynamical equations are

$$\begin{aligned} \frac{dk_{in}^i(t)}{dt} &= \frac{k_{in}^i}{\sum_j k_{in}^j} + mp_{k_{in}^i}^+, \\ \frac{dk_{out}^i(t)}{dt} &= \frac{k_{out}^i}{\sum_j k_{out}^j} + mp_{k_{out}^i}^+, \end{aligned} \quad (7)$$

where the first term in the equations comes from the newly added node in a time step. The difference of the two equations gives

$$\frac{d[k_{in}^i(t) - k_{out}^i(t)]}{dt} = mp_{k_{in}^i}^+ - mp_{k_{out}^i}^+ = \frac{m(1-p)(k_{in}^i - k_{out}^i)}{(2 + m + mp)t}, \quad (8)$$

where Eqs. (5) and (6) have been used. Let t_i be the time that the node i is introduced, i. e., $k_{in}^i(t_i) = k_{out}^i(t_i) = 1$. It follows from Eq. (8) that $k_{in}^i(t) = k_{out}^i(t)$ at any time t . Although the expected value of the difference between indegrees and outdegrees of a node does not grow over time mathematically, the difference does exist in a particular realization of the model in simulations. Eq. (7) and the initial condition $k_{in}^i(t_i) = k_{out}^i(t_i) = 1$ gives

$$k_{in}^i(t) = k_{out}^i(t) = \left(\frac{t}{t_i}\right)^\beta, \quad (9)$$

where $\beta = (1 + m + mp)/(2 + m + mp)$. The indegree and outdegree of the nodes both grow over time in the same functional form, with older nodes having higher indegrees and outdegrees.

Let $N_{k_{in}}(t)$ and $N_{k_{out}}(t)$ be the number of nodes with expected indegree k_{in} and outdegree k_{out} at the time step t , respectively. The rate equation of $N_{k_{in}}(t)$ is then given by

$$\frac{dN_{k_{in}}(t)}{dt} = \frac{k_{in} - 1}{\sum_j k_{in}^j} N_{k_{in}-1} - \frac{k_{in}}{\sum_j k_{in}^j} N_{k_{in}} + mp_{k_{in}-1}^+ N_{k_{in}-1} - mp_{k_{in}}^+ N_{k_{in}} + \delta_{k_{in},1}. \quad (10)$$

The first and third terms on the right-hand side account for the increase of $N_{k_{in}}(t)$ due to the external reciprocation and internal evolution, respectively; and the second and fourth terms account for the decrease due to the processes. The last term accounts for the introduction of a new node with indegree $k_{in}=1$ at time t . Eq. (10) is valid for all $k_{in} \geq 1$.

After many steps t , there are $N = m_0 + t \approx t$ nodes in the network. In the asymptotic limit, we substitute $N_{k_{in}}(t) = tP(k_{in})$, where $P(k_{in})$ is the indegree distribution [59], and $\sum_j k_{in}^j = (2 + m + mp)t$ into Eq. (10) to obtain the simple recursive relation

$$[2 + m + mp + (1 + m + mp)k_{in}]P(k_{in}) = (1 + m + mp)(k_{in} - 1)P(k_{in} - 1) + (2 + m + mp)\delta_{k_{in},1}. \quad (11)$$

Using the initial condition that $k_{in} = 1$ at the time that a node was introduced, the solution of Eq.(11) is

$$P(k_{in}) = A \frac{\Gamma(k_{in})}{\Gamma(k_{in} + 2 + \frac{1}{1+m+mp})}, \quad (12)$$

where $A = \frac{2+m+mp}{3+2m+2mp} \Gamma(\frac{1}{1+m+mp} + 3)$ and Γ is the Euler gamma function. Using the asymptotic form $\Gamma(x + \lambda) \rightarrow x^\lambda$ as $x \rightarrow \infty$, we can extract the scaling form

$$P(k_{in}) \approx A k_{in}^{-(2+\frac{1}{1+m+mp})}. \quad (13)$$

Similarly, the rate equation of $N_{k_{out}}(t)$ is given by

$$\frac{dN_{k_{out}}(t)}{dt} = \frac{k_{in} - 1}{\sum_j k_{in}^j} N_{k_{out}-1} - \frac{k_{in}}{\sum_j k_{in}^j} N_{k_{out}} + mp_{k_{out}-1}^+ N_{k_{out}-1} - mp_{k_{out}}^+ N_{k_{out}} + \delta_{k_{out},1}. \quad (14)$$

The first (second) and third (fourth) terms on the right-hand side account for the increase (decrease) in $N_{k_{out}}$ due to the external reciprocation and internal evolution, respectively; and the last term accounts for the introduction of a new node with $k_{out} = 1$ at time t . Substituting $N_{k_{out}}(t) = tP(k_{out})$, where $P(k_{out})$ is the outdegree distribution, and $\sum_j k_{in}^j = (2 + m + mp)t$ into Eq. (14), the recursive relation for $P(k_{out})$ is

$$[2 + m + mp + (1 + m + mp)k_{out} + 1]P(k_{out}) = (1 + m + mp)(k_{out} - 1)P(k_{out} - 1) + (2 + m + mp)\delta_{k_{out},1}, \quad (15)$$

which is identical to Eq. (11) for $P(k_{in})$. It follows that

$$P(k_{out}) \approx A k_{out}^{-(2+\frac{1}{1+m+mp})}. \quad (16)$$

The results show that the expected indegree and outdegree grow over time following the same functional form of Eq. (9), and the indegree and outdegree distributions follow the same scaling law with an exponent

$$\gamma = 2 + \frac{1}{1 + m + mp}. \quad (17)$$

Next, we consider the reciprocal degree distribution $P(k_r)$. For a node i with $k_{in}^i = k_{out}^i$, k_r^i satisfies the dynamical equation

$$\frac{dk_r^i(t)}{dt} = \frac{k_{in}^i(t)}{(2 + m + mp)t} + \frac{mpk_{in}^i(t) + mpk_{out}^i(t)}{(2 + m + mp)t}. \quad (18)$$

Substituting Eq. (9) into Eq. (18) and using the initial condition that $k_r^i(t_i) = 1$ at the time that node i was introduced, the solution to Eq. (18) is

$$k_r^i(t) = \frac{2mp + 1}{1 + m + mp} [k_{in}^i(t) - 1] + 1. \quad (19)$$

For large k_{in}^i , we have

$$k_r^i \sim \frac{2mp + 1}{1 + m + mp} k_{in}^i. \quad (20)$$

Using $P(k_r)dk_r = P(k_{in})dk_{in}$, the distribution $P(k_r)$ follows

$$P(k_r) \sim k_r^{-\gamma}, \quad (21)$$

where γ is given by Eq. (17) as for the indegree and outdegree distributions.

Furthermore, we analyze the degree correlations between connected nodes by the rate equation approach. Let $N_{k_{in}}^{l_{out}}$ be the number of links that originate from a node with an expected outdegree l_{out} to a node with an expected indegree k_{in} [60]. Generally, $P_{k_{in}}^{l_{out}}$ is defined for $k_{in} \geq 1$ and $l_{out} \geq 2$. The quantity $N_{k_{in}}^{l_{out}}(t)$ evolves according to

$$\begin{aligned} \frac{dN_{k_{in}}^{l_{out}}(t)}{dt} = & \frac{(k_{in} - 1)N_{k_{in}-1}^{l_{out}} - k_{in}N_{k_{in}}^{l_{out}}}{\sum k_{in}N_{k_{in}}} + \frac{(l_{out} - 1)N_{k_{in}}^{l_{out}-1} - l_{out}N_{k_{in}}^{l_{out}}}{\sum k_{in}N_{k_{in}}} + \frac{(l_{out} - 1)N_{l_{out}-1}}{\sum k_{in}N_{k_{in}}} \delta_{1,k_{in}} \\ & + mp_{k_{in}-1}^+ N_{k_{in}-1}^{l_{out}} - mp_{k_{in}}^+ N_{k_{in}}^{l_{out}} + mp_{l_{out}-1}^+ N_{k_{in}}^{l_{out}-1} - mp_{l_{out}}^+ N_{k_{in}}^{l_{out}} \\ & + (m + mp) \frac{(l_{out} - 1)(k_{in} - 1)}{\sum_{x,y,x \notin \Gamma_{in}(y)} k_{out}^x k_{in}^y} N_{l_{out}-1} N_{k_{in}-1} - (m + mp) \frac{l_{out} k_{in}}{\sum_{x,y,x \notin \Gamma_{in}(y)} k_{out}^x k_{in}^y} N_{l_{out}} N_{k_{in}}, \end{aligned} \quad (22)$$

where the first two terms on the right-hand side account for the changes due to the introduction of a new node, including the gains when the new node is connected to a node with indegree $(k_{in} - 1)$ (outdegree $(l_{out} - 1)$) which is already connected to a node with outdegree l_{out} (indegree k_{in}), and the losses when the new node is connected to either end of a link that connects a node with outdegree l_{out} and another node with indegree k_{in} . The third term accounts for the gain in $N_1^{l_{out}}$ due to the addition of the new node. The remaining terms take into account the changes due to the internal evolution process with the introduction of $m + mp$ directed links.

Asymptotically, $N_{k_{in}}^{l_{out}} \rightarrow (2 + m + mp)tP_{k_{in}}^{l_{out}}$, $N_{k_{in}} \rightarrow tP(k_{in})$ and $N_{l_{out}} \rightarrow tP(l_{out})$. Considering $\sum k_{in}N_{k_{in}} = \sum_j k_{in}^j = (2 + m + mp)t$ and $\sum_{x,y,x \notin \Gamma_{in}(y)} k_{out}^x k_{in}^y \approx \sum k_{out}^x \times \sum k_{in}^y = [(2 + m + mp)t]^2$, Eq. (22) gives a recursive relation

$$\begin{aligned} [2 + m + mp + (1 + m + mp)(k_{in} + l_{out})]P_{k_{in}}^{l_{out}} = & (1 + m + mp)[(k_{in} - 1)P_{k_{in}-1}^{l_{out}} + (l_{out} - 1)P_{k_{in}}^{l_{out}-1}] \\ & + \frac{1}{2 + m + mp}(l_{out} - 1)P(l_{out} - 1)\delta_{k_{in},1} \\ & + \frac{m + mp}{(2 + m + mp)^2}[(l_{out} - 1)(k_{in} - 1)P(l_{out} - 1)P(k_{in} - 1) \\ & - l_{out}k_{in}P(l_{out})P(k_{in})]. \end{aligned} \quad (23)$$

Solving Eq. (23) directly for $P_{k_{in}}^{l_{out}}$ is difficult, however, it is observed that decomposing $P_{k_{in}}^{l_{out}}$ into

$$P_{k_{in}}^{l_{out}} \sim l_{out}P(l_{out})k_{in}P(k_{in}), \quad (24)$$

with $P(l_{out})$ given by Eq. (16) and $P(k_{in})$ given by Eq. (13) satisfies Eq. (23) in the scaling regime, as one can readily show by substituting Eq. (24) into Eq. (23) and taking the limits of $l_{out} \rightarrow \infty$ and $k_{in} \rightarrow \infty$. Eq. (24) implies that there is no degree correlation, a feature that is supported by the empirical results in Figure 7 for ESN over a wide range of degrees (also see Figures S16-S18 of Appendix SI). It also follows from $k_{in}^i = k_{out}^i$ and Eq. (24) that $P_{k_{in}}^{l_{out}} = P_{k_{out}}^{l_{out}} = P_{k_{in}}^{l_{in}} = P_{k_{out}}^{l_{in}}$. Interpreting $P_{k_{in}}^{l_{out}}$ as a joint probability, the lack of degree correlation as expressed in Eq. (24) implies that the conditional probability

$$P(k_{in}|l_{out}) \sim k_{in}P(k_{in}), \quad (25)$$

which is independent of l_{out} . For a node i with large $k_{in}^i = k_{out}^i$, the average nearest neighbor function can be calculated as

$$k_{in}^{nn}(k_{out}) = \sum_{k'_{in}} k'_{in} P(k'_{in}|k_{out}) \sim \sum_{k'_{in}} k'^2_{in} P(k'_{in}), \quad (26)$$

which is also independent of k_{out} . This is consistent with the behavior of $k_{in}^{nn}(k_{out})$ in ESN, as shown in Figure 7.

The number of FB loops can be formally written as [61]

$$n_{FB} = \frac{C_{k_{in}}^1 C_{k_{out}}^1}{2} \sum_{k'_{in}, k''_{in}} P(k'_{in}|k_{out}) P(k''_{out}|k_{in}) P_{k'_{in}}^{k''_{out}}, \quad (27)$$

where $P_{k'_{in}}^{k''_{out}}$ is the probability that a link connects a node with outdegree k''_{out} to a node with indegree k'_{in} . The lack of degree correlations makes the summations independent of k_{in} and k_{out} , and thus n_{FB} scales as

$$n_{FB} \sim k_{in}^2. \quad (28)$$

Similarly, the numbers of four closed triples n_{Δ} at a node with large indegree and outdegree follow the scaling behavior $n_{\Delta} \sim k_{in}^2$ or $n_{\Delta} \sim k_{out}^2$. Combining $n_{\Delta} \sim k_{in}^2$ with Eq. (13) ($P(k_{in}) \sim k_{in}^{\gamma_{in}}$), the distributions of four closed triples have the same scaling behavior as follows:

$$P(n_{\Delta}) \sim n_{\Delta}^{\gamma_{\Delta}}, \quad (29)$$

where the exponent γ_{Δ} can be readily found by using $P(n_{\Delta})dn_{\Delta} = P(k_{in})dk_{in}$ to be

$$\gamma_{\Delta} = \frac{3}{2} + \frac{1}{2(1+m+mp)}. \quad (30)$$

The exponent γ_{Δ} is determined by the parameters m and p and it falls into the range $(1.5, 2]$.

Simulation Results

We also carried out numerical simulations to study the structural properties of the model and compared results with data of real online social networks. The activity m and reciprocation probability p are two important parameters of the model. They determine the reciprocity $r = (1+mp)/(1+m)$ and mean indegree $\langle k \rangle = 2+m+mp$ of simulated networks. In order to compare results with real online social networks, we take three parameters from real data, namely the number of nodes N , the reciprocity r and the mean indegree (outdegree) $\langle k \rangle$, and determine the parameter m and p in the model through

$$\begin{aligned} m &= \frac{\langle k \rangle}{1+r} - 1; \\ p &= \frac{\langle k \rangle r - r - 1}{\langle k \rangle - r - 1}. \end{aligned} \quad (31)$$

Taking ESN as an example, we have $\langle k \rangle \approx 6.7$, $r \approx 0.25$, and $N = 75879$. The model parameters are then fixed at $m \approx 4.34$ and $p \approx 0.08$ according to Eq.(31). With the values of m and p , a network of $N = 75879$ nodes is simulated. For a non-integer value of m , it is implemented in a probabilistic way. For ESN with $m = 4.34$, for example, the initiation of the fifth new directed link through the internal evolution process is implemented with a probability 0.34 after establishing four new directed links in every time step. The structural properties of the simulated network are analyzed for each of the quantities studied for the real data. Results are shown in Figures 2-9 as red circles for comparison (also see Figures S1-S24 of Appendix SI). The model basically reproduces the key properties of ESN.

For the indegree and outdegree distributions (see Figure 2) and the reciprocal degree distribution (see Figure 4), the simulation results also show similar scaling law, with the exponents $\gamma_{in} \approx 1.95$, $\gamma_{out} \approx 1.96$ and $\gamma_r \approx 2.1$ determined by the maximum likelihood estimation [54, 55] (see Table S1 of Appendix SI for more detail). These values are slightly larger than the corresponding values of the exponents in ESN. According to Eqs. (17), (21) and (31), these exponents are equal and the theoretical value is $\gamma = 2+1/(\langle k \rangle - 1) \approx 2.17$. Note that the rate equation analysis assumes an infinite system. The difference

between the simulated results and the theoretical value comes from the finite size of simulated network, as well as the approximations made in getting at the values of the exponent. The indegree and outdegree distributions of simulated network are in reasonable agreement with the empirical results of ESN. The model, however, gives a reciprocal degree distribution smaller than the ESN empirical results over a wide range of k_r . This discrepancy implies that there are some network growing mechanisms in ESN that are not included in the model, *e.g.*, different reciprocation probabilities for different nodes [62]. This, together with a possibly very weak degree correlation in Figure 7 that we ignored, may be the reason for the simulation results in Figures 3 and 5 to be bigger than the empirical values for large in/outdegrees, and for the small differences in the tails in Figures 2 and 4 [63, 64].

For the distributions of the four closed triples, the distributions from simulations follow a power-law behavior with almost the same exponent (see Figure 6), where $\gamma_{FB} \approx 1.47$, $\gamma_{FF_a} \approx 1.46$, $\gamma_{FF_b} \approx 1.46$ and $\gamma_{FF_c} \approx 1.46$ as determined by the maximum likelihood estimation. These values are slightly larger than the exponents found in ESN. Theoretically, $\gamma_\Delta = 3/2 + 1/[2(\langle k \rangle - 1)] \approx 1.58$ according to Eqs. (30) and (31). We note that the theoretical values of both γ and γ_Δ depend only on the mean indegree $\langle k \rangle$, which in turn is determined by the two model parameters m and p . Figure 10 shows the values of all the γ -exponents of the distributions for the four online social networks and the corresponding simulated networks, which are determined by the maximum likelihood estimation.

The two parameters m and p affect the reciprocal degree of nodes k_r through Eq.(20). Substituting Eq. (31) into Eq. (20), we have $k_r \sim (2\langle k \rangle r - r - 1)k_{in}/[(1+r)(\langle k \rangle - 1)] \approx 0.3k_{in}$ for ESN. The reciprocal degree k_r of a node and its k_{in} are related by a factor depending on the two global parameters $\langle k \rangle$ and r . This linear relationship between k_r and k_{in} (k_{out}) with a slope 0.3 is observed in simulation results, as shown in Figure 5, but the ESN data show a faster increase of k_r with k_{in} and k_{out} . When the network has a larger reciprocity, such as $r \approx 0.73$ for *Slashdot*, $r \approx 0.45$ for *Flicker*, and $r \approx 0.65$ for *YouTube*, a better agreement is observed (see Figures S10-S12 of Appendix SI). Despite some small differences in the tail in Figures 8 and 9, which may be caused by local proximity bias in link creation [5], simulation results for the dependence of the number of closed triples with k_{in} and k_{out} are basically in accordance with empirical results.

More comparison of results between the model and large-scale online social networks are given in Appendix SI (see Figures S1-S24). The results further support the notions that the two mechanisms incorporated in our model provide a potential explanation of the local and mesoscale structures in these online social networks.

Discussion

With the advancement in information technology, online social systems become an increasingly important part of modern life. It is, therefore, of great significance to study the structures and dynamics of these systems. In this study, we focused on the local scale, mesoscale and macroscale structural properties of online social networks, especially the influence of properties on the local scale and macroscale on the mesoscale structures. We analyzed the data and extracted the local scale and macroscale structural properties of four large-scale online social networks. It was found that the indegree and outdegree distributions follow a similar scaling law, which follows from the fact that $k_{in} \approx k_{out}$ for most of the nodes. It implies that there is a preferential attachment mechanism in which the product $k_{out}^i k_{in}^j$ is important in the establishment of links during the evolution of online social networks. In addition, the very large reciprocity r observed in these networks suggests the existence of a reciprocation mechanism in online social networks. The reciprocal degree distribution also shows a similar exponent as that of the indegree distribution due to the roughly linear relationship between the reciprocal degree k_r and the indegree k_{in} of nodes (*i.e.*, $k_r \sim k_{in}$), which in turn implies a fixed probability of reciprocal links between connected nodes. In the mesoscale, the close-knit friendship structures are determined by both local scale (*i.e.*, indegree and outdegree $k_{in} \approx k_{out}$) and macroscale (*i.e.*, mean in/outdegree $\langle k \rangle$) structural

properties. For a node with large $k_{in} \approx k_{out}$, the numbers of the four closed triples show the same scaling behavior: $n_{FB} \sim k_{in}^2$ and $n_{FF} \sim k_{in}^2$, as a result of the negligible degree correlations in these networks. For all nodes, the distributions of these closed triples also follow a similar scaling law. Despite the numbers of the three feedforward loops are equal, their distributions look somewhat different in detail.

To reproduce the empirical features, we proposed and studied a simple directed network model incorporating an external reciprocation process and an internal evolution process. The two parameters in the model are the activity m and the reciprocation probability p . They can be inferred from the reciprocity r and mean indegree $\langle k \rangle$ of real online social networks according to Eq.(31), so as to ensure that the simulated network and the real network have the same reciprocity and mean indegree. Analytically, we derived the structural properties in the local-scale and mesoscale. The results show that the exponents characterizing the distributions of indegree, outdegree, reciprocal degree and four closed triples depend only on the mean indegree $\langle k \rangle$, *i.e.*, $\gamma = 2 + 1/(\langle k \rangle - 1)$ and $\gamma_{\Delta} = 3/2 + 1/[2(\langle k \rangle - 1)]$. In addition, the mean indegree $\langle k \rangle$ and the reciprocity r together determine the ratio of the reciprocal degree to the directed in/outdegree, *i.e.*, $k_r \sim (2\langle k \rangle r - r - 1)k_{in}/[(\langle k \rangle - 1)(1 + r)]$. The expected indegree and outdegree of nodes in the model grow as the same function of the time that the nodes are introduced, with very old nodes having very high indegrees and outdegrees. This phenomenon, coupled with an essentially fixed rate of reciprocation, reproduces almost all the properties of the online social networks studied here.

The mesoscale structural properties reported in our work help us understand the interplay between structural properties on different scales in online social networks. More specifically, the mesoscale structures in these online social networks are determined by global parameters as well as by local distributions. This provides a useful perspective of future studies in social network analysis. Our work also provides a better understanding of the evolution of online social networks, especially the emergence of close-knit friendship structures with a scaling behavior in their distributions. The two processes (reciprocation and preferential attachment) provide a possible explanation of the mechanisms underlying the local scale and mesoscale structural properties of online social networks. The former reflects that users often respond to a new incoming link by quickly establishing a reversed link. The latter means that a well-known user with a large k_{in} is more likely to attract new connections and an active user with a large k_{out} is more likely to create new connections. Our model may also be applied to other growing directed networks in which the indegree and outdegree distributions show a similar scaling behavior and the reciprocation mechanism is valid. However, the model is not applicable to the symmetric online social networks that lack the power-law degree distributions [1–3] (*e.g.*, Facebook), and to the WWW [33] and Wikipedia [56] as the indegree and outdegree distributions in these systems carry different exponents and the reciprocation mechanism is absent. Similarly, it does not apply to the citation network as a paper can only cite published papers, but not vice versa.

Although simulated results of our model basically reproduced the structural properties of the online social networks at different scales, the differences in the exponents characterizing the distributions and in the tails of the distributions in real online social networks (*e.g.*, Figures 4, 5, 8, 9) imply that there exist other factors, such as individual users of different reciprocation probabilities and local proximity bias, that are ignored in the model. These factors are good ingredients for future work. It is also important to study the emergence of communities in online social networks. The present work also forms the basis for the understanding of the impact of mesoscale structural properties on dynamical processes on online social networks, such as information diffusion, opinions formation, and cooperation evolution. An interesting problem for future work is to investigate whether the model can be applied to offline real social networks. Such a work would help reveal the difference between online and offline social networks.

Supporting information

Appendix SI Appendix to the manuscript.
(PDF)

Table S1 The exponents of various distributions obtained by power-law fits of real online social networks and the simulated network based on the model using the maximum likelihood estimation. x_{min} is the lower bound of the range for fitting a power-law distribution, γ is the corresponding exponent and KS is the goodness-of-fit value based on the Kolmogorov-Smirnov statistic.
(PDF)

Table S2 Pearson correlation coefficient. $r(in, in)$ quantifies the tendency of nodes with a high indegree to be connected to another node with a high indegree. The other quantities carry a similar interpretation.
(PDF)

Figure S1 Indegree (a) and outdegree (b) distributions of the Slashdot social network (black squares) and simulation results (red circles) based on the model. The dashed lines in both panels have a slope -2.1 as the analytic results in Eqs. (17) and (31) suggested. The simulated network is generated by the model with the parameters $N = 82168$, $m \approx 5.14$ and $p \approx 0.67$ as determined by the mean degree $\langle k \rangle$ and reciprocity of the Slashdot social network. Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.
(PDF)

Figure S2 Indegree (a) and outdegree (b) distributions of the Flickr social network (black squares) and simulation results (red circles) based on the model. The dashed lines in both panels have a slope -2.08 as the analytic results in Eqs. (17) and (31) suggested. The simulated network is generated by the model with the parameters $N = 100000$, $m \approx 8.07$ and $p \approx 0.39$ as determined by the mean degree $\langle k \rangle$ and reciprocity of the Flickr social network. Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.
(PDF)

Figure S3 Indegree (a) and outdegree (b) distributions of the YouTube social network (black squares) and simulation results (red circles) based on the model. The dashed lines in both panels have a slope -2.3 as the analytic results in Eqs. (17) and (31) suggested. The simulated network is generated by the model with the parameters $N = 100000$, $m \approx 4.34$ and $p \approx 0.08$ as determined by the mean degree $\langle k \rangle$ and reciprocity of the YouTube social network. Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.
(PDF)

Figure S4 Relationship between the indegree and the outdegree of nodes in the Slashdot social network and the model. Results of the Slashdot social network (black squares) and simulation results (red circles) based on the model are shown. The blue dash line represents the relation function $k_{in} = k_{out}$. Data points are averages over the logarithmic bins of the indegree k_{in} .
(PDF)

Figure S5 Relationship between the indegree and the outdegree of nodes in the Flickr social network and the model. Results of the Flickr social network (black squares) and simulation results (red circles) based on the model are shown. The blue dash line represents the relation function $k_{in} = k_{out}$. Data points are averages over the logarithmic bins of the indegree k_{in} .
(PDF)

Figure S6 Relationship between the indegree and the outdegree of nodes in the YouTube social network and the model. Results of the YouTube social network (black squares) and simulation results (red circles) based on the model are shown. The blue dash line represents the relation function $k_{in} = k_{out}$. Data points are averages over the logarithmic bins of the indegree k_{in} .
(PDF)

Figure S7 Reciprocal degree distributions of the Slashdot social network and the model. Results of the Slashdot social network (black squares) and simulation results (red circles) based on the model are shown. Analytic treatment (see Eqs. (17) and (31)) suggests a scaling behavior with an exponent -2.1 , as shown by the dash line. Data points are averages over the logarithmic bins of the

reciprocal degree k_r .

(PDF)

Figure S8 Reciprocal degree distributions of the Flickr social network and the model. Results of the Flickr social network (black squares) and simulation results (red circles) based on the model are shown. Analytic treatment (see Eqs. (17) and (31)) suggests a scaling behavior with an exponent -2.08 , as shown by the dash line. Data points are averages over the logarithmic bins of the reciprocal degree k_r .

(PDF)

Figure S9 Reciprocal degree distributions of the YouTube social network and the model. Results of the YouTube social network (black squares) and simulation results (red circles) based on the model are shown. Analytic treatment (see Eqs. (17) and (31)) suggests a scaling behavior with an exponent -2.3 , as shown by the dash line. Data points are averages over the logarithmic bins of the reciprocal degree k_r .

(PDF)

Figure S10 Mean reciprocal degree of nodes with (a) the same indegree and (b) the same outdegree in the Slashdot social network and in the model. Results of the Slashdot social network (black squares) and simulation results (red circles) based on the model are shown in a log-log scale in the main panels. Analytic treatment suggests that $\langle k_r \rangle$ is linearly dependent on k_{in} and k_{out} , and the blue dash lines of slope 1 show its dependence. The inset in each panel shows the results in a linear scale and the dash line has a slope of 0.82, as given by Eqs. (20) and (31). Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

(PDF)

Figure S11 Mean reciprocal degree of nodes with (a) the same indegree and (b) the same outdegree in the Flickr social network and in the model. Results of the Flickr social network (black squares) and simulation results (red circles) based on the model are shown in a log-log scale in the main panels. Analytic treatment suggests that $\langle k_r \rangle$ is linearly dependent on k_{in} and k_{out} , and the blue dash lines of slope 1 show its dependence. The inset in each panel shows the results in a linear scale and the dash line has a slope of 0.59, as given by Eqs. (20) and (31). Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

(PDF)

Figure S12 Mean reciprocal degree of nodes with (a) the same indegree and (b) the same outdegree in the YouTube social network and in the model. Results of the YouTube social network (black squares) and simulation results (red circles) based on the model are shown in a log-log scale in the main panels. Analytic treatment suggests that $\langle k_r \rangle$ is linearly dependent on k_{in} and k_{out} , and the blue dash lines of slope 1 show its dependence. The inset in each panel shows the results in a linear scale and the dash line has a slope of 0.73, as given by Eqs. (20) and (31). Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

(PDF)

Figure S13 Distributions of four basic closed triples in the slashdot social network and the model. Distributions of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops in the Slashdot social network (black squares) and in the simulated network based on the model (red circles). Analytic treatment (see Eqs. (30) and (31)) suggests a scaling behavior with an exponent -1.55 , as shown by the dash lines. Data points are averages over the logarithmic bins of the n_{FB} , n_{FF_a} , n_{FF_b} and n_{FF_c} , respectively.

(PDF)

Figure S14 Distributions of four basic closed triples in the Flickr social network and the model. Distributions of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops in the Flickr social network (black squares) and in the simulated network based on the model (red circles). Analytic treatment (see Eqs. (30) and (31)) suggests a scaling behavior with an exponent -1.54 , as shown by the dash lines. Data points are averages over the logarithmic bins of the n_{FB} , n_{FF_a} , n_{FF_b} and

n_{FFc} , respectively.

(PDF)

Figure S15 Distributions of four basic closed triples in the YouTube social network and the model. Distributions of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops in the YouTube social network (black squares) and in the simulated network based on the model (red circles). Analytic treatment (see Eqs. (30) and (31)) suggests a scaling behavior with an exponent -1.65 , as shown by the dash lines. Data points are averages over the logarithmic bins of the n_{FB} , n_{FFa} , n_{FFb} and n_{FFc} , respectively.

(PDF)

Figure S16 Degree correlations in the Slashdot social network and the model. Results of degree correlations as measured by four quantities corresponding to the average nearest neighbor degree $\langle k_{in}^{nn}(k_{in}) \rangle$ (squares), $\langle k_{out}^{nn}(k_{in}) \rangle$ (circles), $\langle k_{out}^{nn}(k_{out}) \rangle$ (triangles), and $\langle k_{in}^{nn}(k_{out}) \rangle$ (inverted triangles) for (a) Slashdot social network and (b) simulated network based on the model. Data points are averages over the logarithmic bins of the indegree k_{in} or outdegree k_{out} .

(PDF)

Figure S17 Degree correlations in the Flickr social network and the model. Results of degree correlations as measured by four quantities corresponding to the average nearest neighbor degree $\langle k_{in}^{nn}(k_{in}) \rangle$ (squares), $\langle k_{out}^{nn}(k_{in}) \rangle$ (circles), $\langle k_{out}^{nn}(k_{out}) \rangle$ (triangles), and $\langle k_{in}^{nn}(k_{out}) \rangle$ (inverted triangles) for (a) Flickr social network and (b) simulated network based on the model. Data points are averages over the logarithmic bins of the indegree k_{in} or outdegree k_{out} .

(PDF)

Figure S18 Degree correlations in the YouTube social network and the model. Results of degree correlations as measured by four quantities corresponding to the average nearest neighbor degree $\langle k_{in}^{nn}(k_{in}) \rangle$ (squares), $\langle k_{out}^{nn}(k_{in}) \rangle$ (circles), $\langle k_{out}^{nn}(k_{out}) \rangle$ (triangles), and $\langle k_{in}^{nn}(k_{out}) \rangle$ (inverted triangles) for (a) YouTube social network and (b) simulated network based on the model. Data points are averages over the logarithmic bins of the indegree k_{in} or outdegree k_{out} .

(PDF)

Figure S19 Mean number of the four closed triples for nodes with the same indegree in the Slashdot social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same indegree are shown for the Slashdot social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the indegree k_{in} .

(PDF)

Figure S20 Mean number of the four closed triples for nodes with the same indegree in the Flickr social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same indegree are shown for the Flickr social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the indegree k_{in} .

(PDF)

Figure S21 Mean number of the four closed triples for nodes with the same indegree in the YouTube social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same indegree are shown for the YouTube social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the indegree k_{in} .

(PDF)

Figure S22 Mean number of the four closed triples for nodes with the same outdegree in the

Slashdot social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same outdegree are shown for the Slashdot social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the outdegree k_{out} .

(PDF)

Figure S23 Mean number of the four closed triples for nodes with the same outdegree in the Flickr social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same outdegree are shown for the Flickr social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the outdegree k_{out} .

(PDF)

Figure S24 Mean number of the four closed triples for nodes with the same outdegree in the YouTube social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same outdegree are shown for the YouTube social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the outdegree k_{out} .

(PDF)

Acknowledgments

We acknowledge Tao Zhou and Ying Liu for valuable discussions.

References

1. Corten R (2012) Composition and structure of a large online social network in the netherlands. PLoS ONE 7: e34760.
2. Gjoka M BCTMA Kurant M (2010) Walking in facebook: A case study of unbiased sampling of osns. In: Proceedings of IEEE INFOCOM. ACM.
3. Ahn YY, Han S, Kwak H, Moon S, Jeong H (2007) Analysis of topological characteristics of huge online social networking services. In: Proceedings of the 16th international conference on World Wide Web. ACM, pp. 835–844.
4. Hu HB, Wang XF (2009) Evolution of a large online social network. Phys Lett A 373: 1105–1110.
5. Mislove A, Koppula HS, Gummadi KP, Druschel P, Bhattacharjee B (2008) Growth of the flickr social network. In: Proceedings of the first workshop on Online social networks. ACM, pp. 25–30.
6. Benevenuto F, Duarte F, Rodrigues T, Almeida VAF, Almeida JM, et al. (2008) Understanding video interactions in youtube. In: Proceeding of the 16th ACM international conference on Multimedia. ACM, pp. 761–764.
7. Gómez V, Kaltenbrunner A, López V (2008) Statistical analysis of the social network and discussion threads in slashdot. In: Proceeding of the 17th international conference on World Wide Web. ACM, pp. 645–654.
8. Scott J (1988) Social network analysis. Sociology 22: 109–127.

9. Scott J (2000) Social network analysis: a handbook. Sage.
10. Albert R, Barabási AL (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74: 47-97.
11. Dorogovtsev SN, Mendes JFF (2002) Evolution of networks. *Adv Phys* 51: 1079–1187.
12. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45: 167–256.
13. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) Complex networks: Structure and dynamics. *Phys Rep* 424: 175–308.
14. Costa LF, Rodrigues FA, Travieso G, Boas PRV (2007) Characterization of complex networks: A survey of measurements. *Adv Phys* 56: 167–242.
15. Dorogovtsev SN, Goltsev AV, Mendes JFF (2008) Critical phenomena in complex networks. *Rev Mod Phys* 80: 1275-1335.
16. Barthélemy M (2011) Spatial networks. *Phys Rep* 499: 1–101.
17. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. *Rev Mod Phys* 81: 591-646.
18. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329: 1194-1197.
19. Lü L, Zhang YC, Yeung CH, Zhou T (2011) Leaders in social networks, the delicious case. *PloS ONE* 6: e21202.
20. Zhou T, Medo M, Cimini G, Zhang ZK, Zhang YC (2011) Emergence of scale-free leadership structure in social recommender systems. *PloS ONE* 6: e20648.
21. Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. *Nature* 437: 1291–1298.
22. Brin S, Page L (1999) The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* 30: 107–117.
23. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. *JACM* 46: 604–632.
24. Newman MEJ (2001) The structure of scientific collaboration networks. *Proc Natl Acad Sci USA* 98: 404-409.
25. Leicht EA, Clarkson G, Shedden K, Newman MEJ (2007) Large-scale structure of time evolving citation networks. *Eur Phys J B* 59: 75–83.
26. Rodgers GJ, Darby-Dowman K (2001) Properties of a growing random directed network. *Eur Phys J B* 23: 267–271.
27. Foster JG, Foster DV, Grassberger P, Paczuski M (2010) Edge direction and the structure of networks. *Proc Natl Acad Sci USA* 107: 10815-10820.
28. Palla G, Farkas IJ, Pollner P, Derenyi I, Vicsek T (2007) Directed network modules. *New J Phys* 9: 186-206.
29. Leicht EA, Newman MEJ (2008) Community structure in directed networks. *Phys Rev Lett* 100: 118703.

30. Kim Y, Son SW, Jeong H (2010) Finding communities in directed networks. *Phys Rev E* 81: 016103.
31. Dorogovtsev SN, Mendes JFF, Samukhin AN (2000) Structure of growing networks with preferential linking. *Phys Rev Lett* 85: 4633–4636.
32. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509–512.
33. Krapivsky PL, Rodgers GJ, Redner S (2001) Degree distributions of growing networks. *Phys Rev Lett* 86: 5401–5404.
34. Almendral JA, Criado R, Leyva I, Buldú JM, Sendiña Nadal I (2010) Announcement: Focus issue on “mesoscales in complex networks”. *CHAOS* 20: 010202.
35. Almendral JA, Criado R, Leyva I, Buldú JM, Sendiña Nadal I (2011) Introduction to focus issue: Mesoscales in complex networks. *CHAOS* 21: 016101.
36. Reichardt J, Alamino R, Saad D (2011) The interplay between microscopic and mesoscopic structures in complex networks. *PloS ONE* 6: e21282.
37. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486: 75–174.
38. Lancichinetti A, Radicchi F, Ramasco JJ, Fortunato S (2011) Finding statistically significant communities in networks. *PloS ONE* 6: e18961.
39. Lancichinetti A, Kivelä M, Saramäki J, Fortunato S (2010) Characterizing the community structure of complex networks. *PloS ONE* 5: e11976.
40. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393: 440–442.
41. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: simple building blocks of complex networks. *Science* 298: 824–827.
42. Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, et al. (2004) Superfamilies of evolved and designed networks. *Science* 303: 1538–1542.
43. Huang CY, Sun CT, Cheng CY, Hsieh JL (2007) Bridge and brick motifs in complex networks. *Physica A* 377: 340–350.
44. Fagiolo G (2007) Clustering in complex directed networks. *Phys Rev E* 76: 026107.
45. Ahnert SE, Fink TMA (2008) Clustering signatures classify directed networks. *Phys Rev E* 78: 036112.
46. Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad of Sci USA* 100: 11980–11985.
47. Mangan S, Itzkovitz S, Zaslaver A, Alon U (2006) The incoherent feed-forward loop accelerates the response-time of the gal system of escherichia coli. *J Mol Biol* 356: 1073–1081.
48. Sousa AO (2005) Consensus formation on a triad scale-free network. *Physica A* 348: 701–710.
49. Ghoneim A, Abbass H, Barlow M (2008) Characterizing game dynamics in two-player strategy games using network motifs. *IEEE Trans Syst Man Cybern Part B: Cybern* 38: 682–690.
50. Hales D, Arteconi S (2008) Motifs in evolving cooperative networks look like protein structure networks. *Networks and Heterogeneous Media* 3: 239–249.

51. Leskovec J, Huttenlocher D, Kleinberg J (2010) Signed networks in social media. In: Proceedings of the 28th international conference on Human factors in computing systems. ACM, pp. 1361–1370.
52. Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, pp. 29–42.
53. Garlaschelli D, Loffredo MI (2004) Patterns of link reciprocity in directed networks. *Phys Rev Lett* 93: 268701.
54. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51: 661–703.
55. Stumpf MPH, Porter MA (2012) Critical truths about power laws. *Science* 335: 665–666.
56. Capocci A, Servedio VDP, Colaiori F, Buriol LS, Donato D, et al. (2006) Preferential attachment in the growth of social networks: The internet encyclopedia wikipedia. *Phys Rev E* 74: 036116.
57. Lü L, Zhou T (2011) Link prediction in complex networks: A survey. *Physica A* 390: 1150–1170.
58. Newman MEJ (2002) Assortative mixing in networks. *Phys Rev Lett* 89: 208701.
59. Krapivsky PL, Redner S, Leyvraz F (2000) Connectivity of growing random networks. *Phys Rev Lett* 85: 4629–4632.
60. Krapivsky PL, Redner S (2001) Organization of growing random networks. *Phys Rev E* 63: 066123.
61. Catanzaro M, Boguñá M, Pastor-Satorras R (2005) Generation of uncorrelated random scale-free networks. *Phys Rev E* 71: 027103.
62. Gallos LK, Rybski D, Liljeros F, Havlin S, Makse HA (2012) How people interact in evolving online affiliation networks. *Phys Rev X* 2: 031014.
63. Wu Y, Zhou C, Xiao J, Kurths J, Schellnhuber HJ (2010) Evidence for a bimodal distribution in human communication. *Proc Natl Acad Sci USA* 107: 18803–18808.
64. Mossa S, Barthélemy M, Eugene Stanley H, Nunes Amaral LA (2002) Truncation of power law behavior in scale-free network models due to information filtering. *Phys Rev Lett* 88: 138701.

Table 1. Basic statistics of the four online social network datasets. Properties of each network: number of users N , number of directed links E , reciprocity r , number of feedback (FB) loop N_{FB} , number of feedforward loops N_{FF_a} . The numbers of the three feedforward loops (FF_a , FF_b , FF_c) are equal, because every FF_a loop from the perspective of the focal node constitutes a FF_b loop and a FF_c loop from the perspective of the another two nodes.

Data sets	Epinions	Slashdot	Flickr	YouTube
N	75,879	82,168	1,715,255	1,138,499
E	508,825	870,161	22,613,980	4,945,382
k_{in}^{max}	3035	2552	16255	25519
k_{out}^{max}	1801	2510	26185	28644
r	0.25	0.73	0.45	0.65
N_{FB}	740,310	899,316	435,829,822	5,320,127
N_{FF_a}	3,586,403	2,881,727	1,667,179,686	16,287,794

Figure Legends

Figure 1. Three possible unclosed triples and four basic closed triples for a focal node (red).

Figure 2. Indegree (a) and outdegree (b) distributions of the Epinions social network (black squares) and simulation results (red circles) based on the model.

Figure 3. Relationship between the indegree and the outdegree of nodes in the Epinions social network and the model.

Figure 4. Reciprocal degree distributions of the Epinions social network and the model.

Figure 5. Mean reciprocal degree of nodes with (a) the same indegree and (b) the same outdegree in the Epinions social network and in the model.

Figure 6. Distributions of four basic closed triples in the Epinions social network and the model.

Figure 7. Degree correlations in the Epinions social network and the model.

Figure 8. Mean number of the four closed triples for nodes with the same indegree in the Epinions social network and the model.

Figure 9. Mean number of the four closed triples for nodes with the same outdegree in the Epinions social network and the model.

Figure 10. Values of the γ -exponents for various distributions.

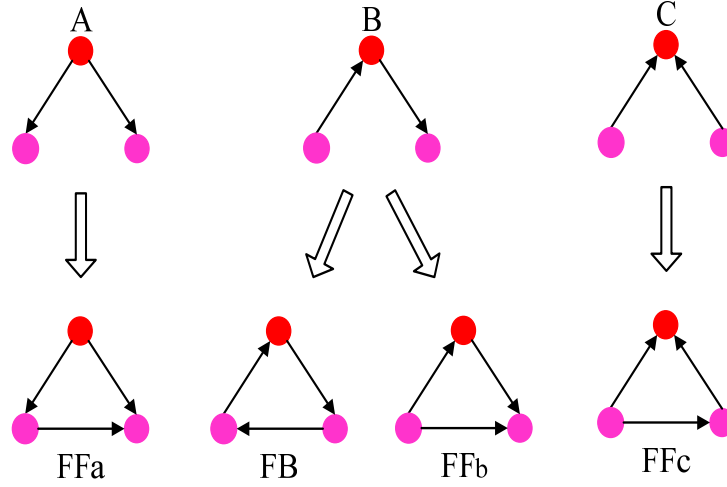


Figure 1. Three possible unclosed triples and four basic closed triples for a focal node (red). The basic closed triples correspond to one feedback (FB) loop and three feedforward (FF) loops. The three feedforward loops differ in the indegree k_{in} of the focal node: $k_{in} = 0$ for FF_a , $k_{in} = 1$ for FF_b and $k_{in} = 2$ for FF_c . The numbers of the three feedforward loops are equal because every FF_a loop from the perspective of the focal node constitutes a FF_b loop and a FF_c loop from the perspective of the another two nodes, but the loops may arise from different growth histories.

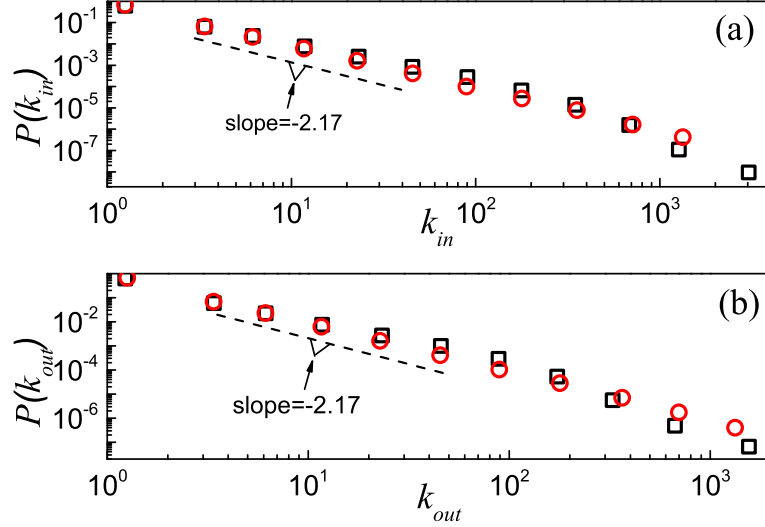


Figure 2. Indegree (a) and outdegree (b) distributions of the Epinions social network (black squares) and simulation results (red circles) based on the model. The dashed lines in both panels have a slope -2.17 as the analytic results in Eqs. (17) and (31) suggested. The simulated network is generated by the model with the parameters $N = 75879$, $m \approx 4.34$ and $p \approx 0.08$, as determined by the mean degree $\langle k \rangle$ and reciprocity r of the Epinions social network. Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

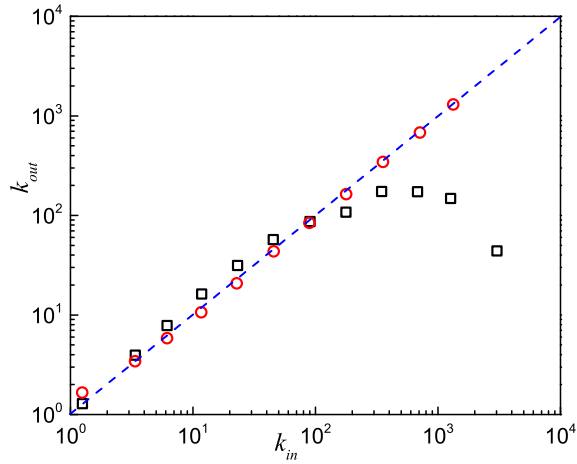


Figure 3. Relationship between the indegree and the outdegree of nodes in the Epinions social network and the model. Results of the Epinions social network (black squares) and simulation results (red circles) based on the model are shown. The blue dash line represents the relation function $k_{in} = k_{out}$. Data points are averages over the logarithmic bins of the indegree k_{in} .

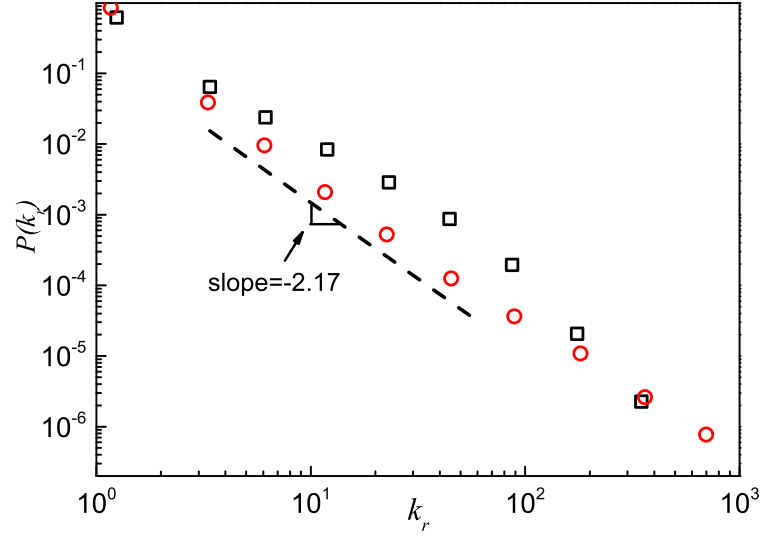


Figure 4. Reciprocal degree distributions of the Epinions social network and the model. Results of the Epinions social network (black squares) and simulation results (red circles) based on the model are shown. Analytic treatment (see Eqs. (17) and (31)) suggests a scaling behavior with an exponent -2.17 , as shown by the dash line. Data points are averages over the logarithmic bins of the reciprocal degree k_r .

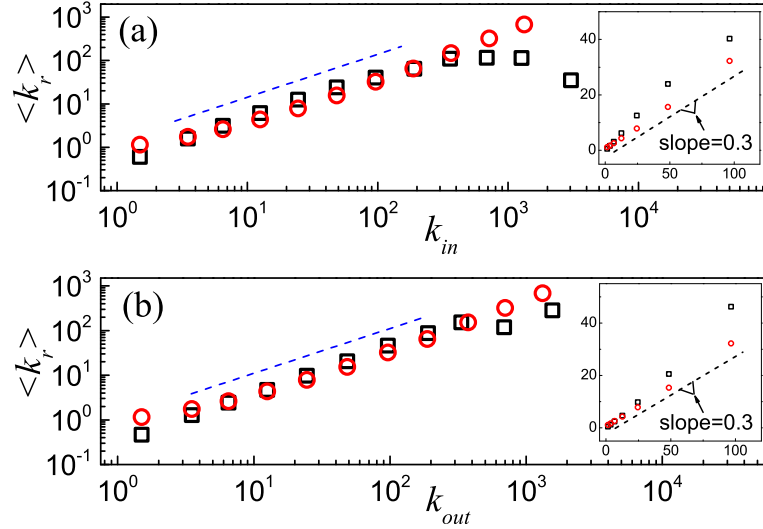


Figure 5. Mean reciprocal degree of nodes with (a) the same indegree and (b) the same outdegree in the Epinions social network and in the model. Results of the Epinions social network (black squares) and simulation results (red circles) based on the model are shown in a log-log scale in the main panels. Analytic treatment suggests that $\langle k_r \rangle$ is linearly dependent on k_{in} and k_{out} , and the blue dash lines of slope 1 show its dependence. The inset in each panel shows the results in a linear scale and the dash line has a slope of 0.3, as given by Eqs. (20) and (31). Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

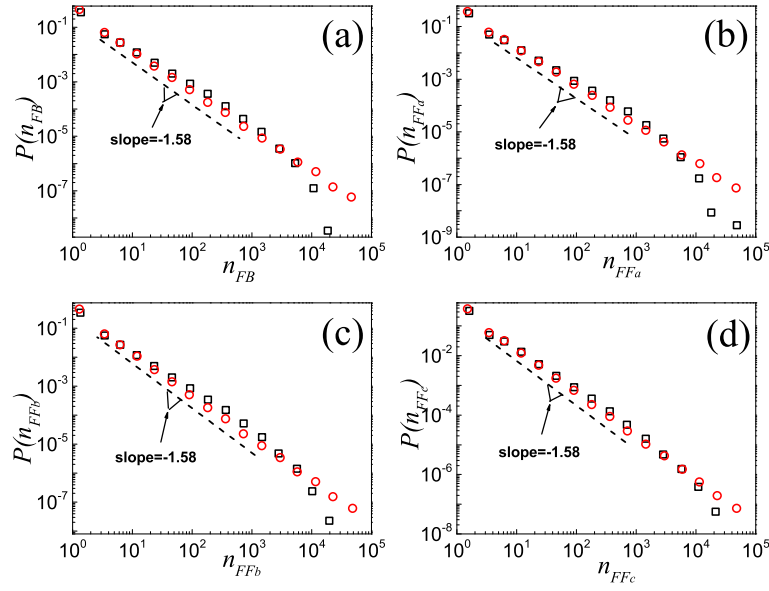


Figure 6. Distributions of four basic closed triples in the Epinions social network and the model. Distributions of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops in the Epinions social network (black squares) and in the simulated network based on the model (red circles). Analytic treatment (see Eqs. (30) and (31)) suggests a scaling behavior with an exponent -1.58 , as shown by the dash lines. Data points are averages over the logarithmic bins of the n_{FB} , n_{FFa} , n_{FFb} and n_{FFc} , respectively.

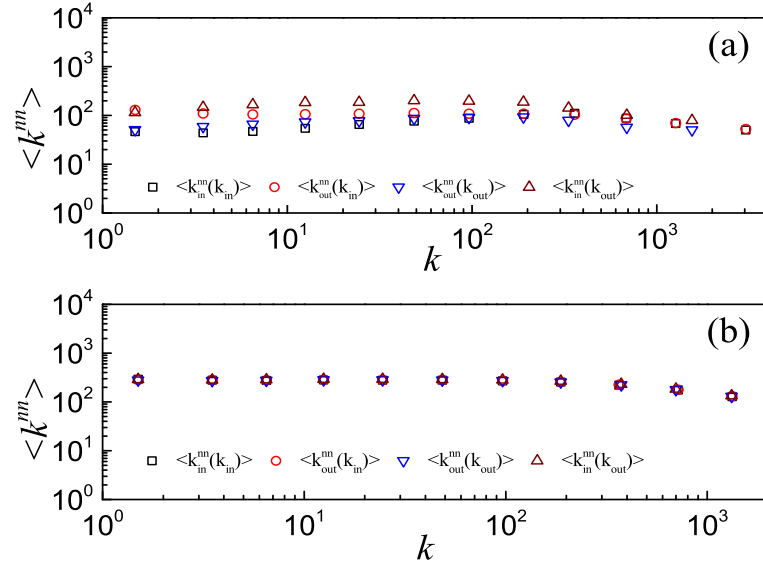


Figure 7. Degree correlations in the Epinions social network and the model. Results of degree correlations as measured by four quantities corresponding to the average nearest neighbor degree $\langle k^{nn}(k_{in}) \rangle$ (squares), $\langle k^{nn}(k_{in}) \rangle$ (circles), $\langle k^{nn}(k_{out}) \rangle$ (triangles), and $\langle k^{nn}(k_{out}) \rangle$ (inverted triangles) for (a) the Epinions social network and (b) simulated network based on the model. Data points are averages over the logarithmic bins of the indegree k_{in} or outdegree k_{out} .

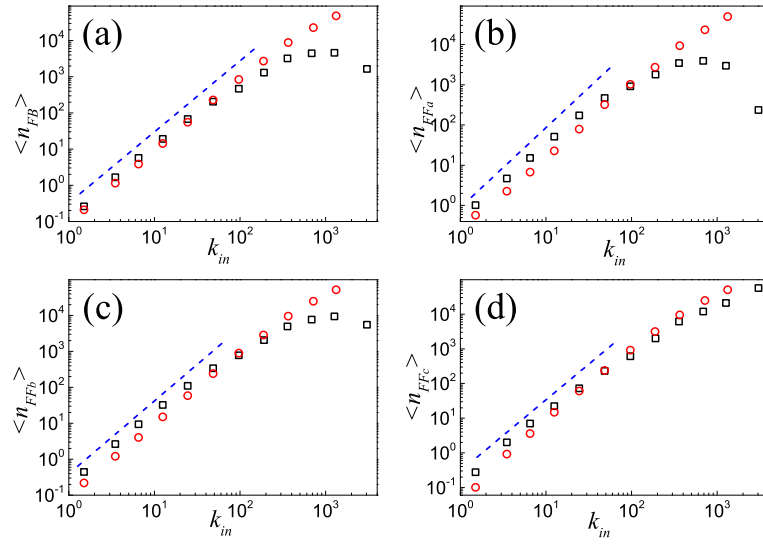


Figure 8. Mean number of the four closed triples for nodes with the same indegree in the Epinions social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same indegree are shown for the Epinions social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash lines. Data points are averages over the logarithmic bins of the indegree k_{in} .

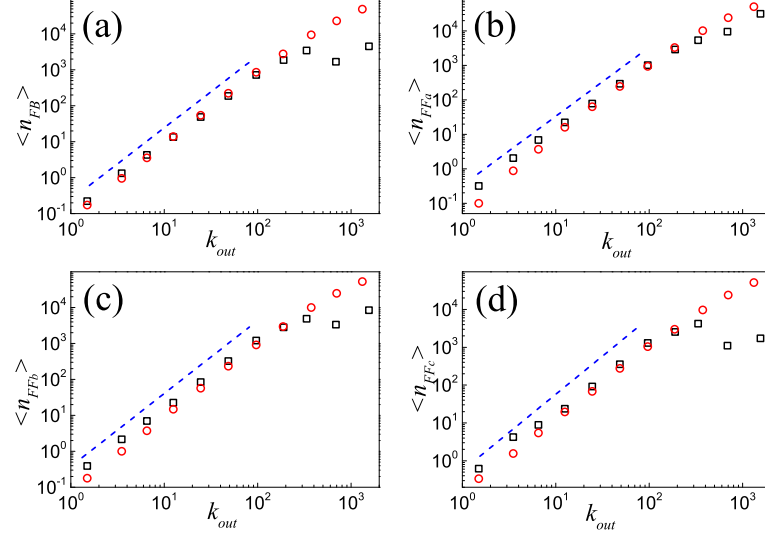


Figure 9. Mean number of the four closed triples for nodes with the same outdegree in the Epinions social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same outdegree are shown for the Epinions social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash lines. Data points are averages over the logarithmic bins of the outdegree k_{out} .

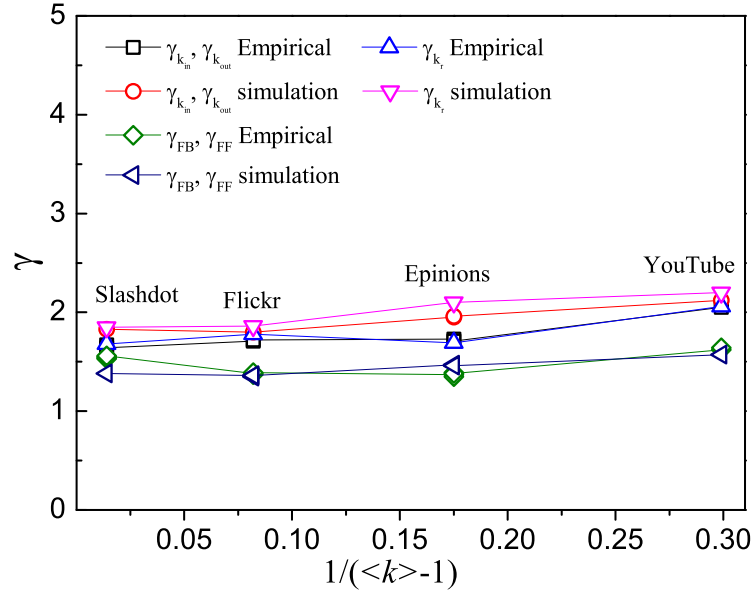


Figure 10. Values of the γ -exponents for various distributions. Values of the γ -exponents for the various distributions as determined by the maximum likelihood estimation against $1/(\langle k \rangle - 1)$ for each of the four large-scale online social networks and the corresponding simulated networks based on the model. The lines are only guides to the eye.

Appendix SI: Emergence of scale-free close-knit friendship structure in online social networks

This Appendix is divided into three sections. In Sec. 1, we give the exponents of power-law fits carried out by the maximum likelihood estimation. In Sec. 2, we present the degree correlation by the Pearson correlation coefficient. In Sec. 3, statistical properties of *Slashdot*, *Flickr* and *YouTube* are presented.

Table S1. The exponents of various distributions obtained by power-law fits of real online social networks and the simulated network based on the model using the maximum likelihood estimation. x_{min} is the lower bound of the range for fitting a power-law distribution, γ is the corresponding exponent and KS is the goodness-of-fit value based on the Kolmogorov-Smirnov statistic.

Distribution	Epinions			model		
	γ	x_{min}	KS	γ	x_{min}	KS
$P(k_{in})$	1.73	2	0.018	1.95	3	0.009
$P(k_{out})$	1.71	2	0.022	1.96	3	0.014
$P(k_r)$	1.69	1	0.01	2.1	4	0.025
$P(n_{FB})$	1.37	2	0.034	1.47	3	0.006
$P(n_{FFa})$	1.39	2	0.04	1.46	5	0.009
$P(n_{FFb})$	1.35	2	0.029	1.46	2	0.008
$P(n_{FFc})$	1.38	3	0.038	1.46	5	0.009
Distribution	Slashdot			model		
	γ	x_{min}	KS	γ	x_{min}	KS
$P(k_{in})$	1.67	2	0.045	1.83	3	0.004
$P(k_{out})$	1.64	2	0.047	1.83	3	0.003
$P(k_r)$	1.68	2	0.047	1.85	2	0.004
$P(n_{FB})$	1.55	6	0.027	1.38	4	0.009
$P(n_{FFa})$	1.53	6	0.029	1.38	4	0.015
$P(n_{FFb})$	1.54	6	0.03	1.38	4	0.008
$P(n_{FFc})$	1.56	6	0.034	1.38	4	0.017
Distribution	Flickr			model		
	γ	x_{min}	KS	γ	x_{min}	KS
$P(k_{in})$	1.71	2	0.015	1.8	4	0.004
$P(k_{out})$	1.72	5	0.03	1.8	4	0.005
$P(k_r)$	1.78	5	0.025	1.86	1	0.006
$P(n_{FB})$	1.38	2	0.029	1.36	4	0.007
$P(n_{FFa})$	1.38	5	0.029	1.35	4	0.01
$P(n_{FFb})$	1.37	2	0.029	1.36	4	0.007
$P(n_{FFc})$	1.39	4	0.033	1.36	5	0.009
Distribution	YouTube			model		
	γ	x_{min}	KS	γ	x_{min}	KS
$P(k_{in})$	2.05	3	0.015	2.12	3	0.013
$P(k_{out})$	2.08	6	0.018	2.12	3	0.015
$P(k_r)$	2.06	6	0.019	2.2	3	0.006
$P(n_{FB})$	1.62	4	0.031	1.57	2	0.01
$P(n_{FFa})$	1.62	4	0.044	1.57	4	0.01
$P(n_{FFb})$	1.62	4	0.029	1.57	4	0.005
$P(n_{FFc})$	1.64	4	0.042	1.57	4	0.01

2. Degree correlation measured by Pearson correlation coefficient

Table S2. Pearson correlation coefficient. $r(in, in)$ quantifies the tendency of nodes with a high indegree to be connected to another node with a high indegree. The other quantities carry a similar interpretation.

Datasets	$r(in, in)$	$r(in, out)$	$r(out, out)$	$r(out, in)$
Epinions	-0.009	0.073	-0.016	-0.053
Slashdot	-0.068	-0.059	-0.064	-0.071
Flickr	0.06	0.055	-0.0025	-0.001
YouTube	-0.03	-0.032	-0.036	-0.035

3. Statistical properties of *Slashdot*, *Flickr* and *YouTube* social networks

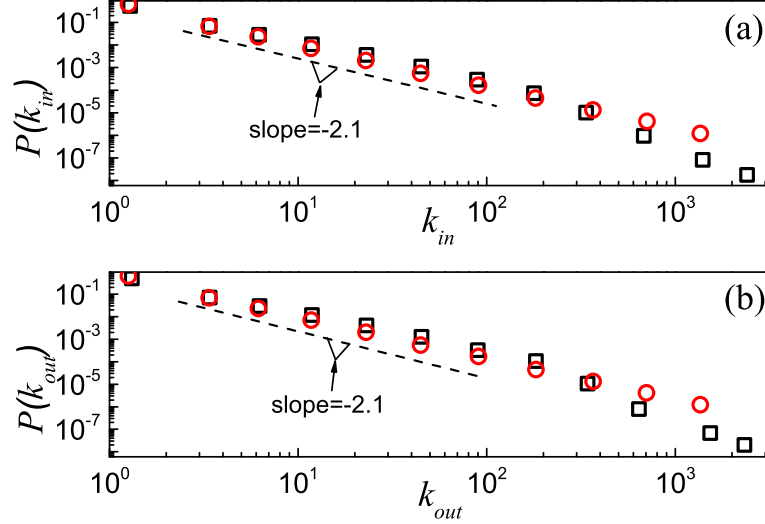


Figure S1. Indegree (a) and outdegree (b) distributions of the Slashdot social network (black squares) and simulation results (red circles) based on the model. The dashed lines in both panels have a slope -2.1 as the analytic results in Eqs. (17) and (31) suggested. The simulated network is generated by the model with the parameters $N = 82168$, $m \approx 5.14$ and $p \approx 0.67$ as determined by the mean degree $\langle k \rangle$ and reciprocity of the Slashdot social network. Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

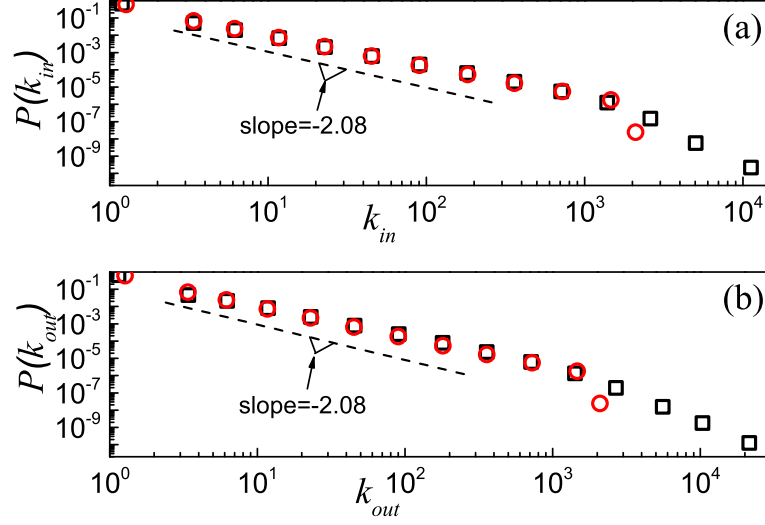


Figure S2. Indegree (a) and outdegree (b) distributions of the Flickr social network (black squares) and simulation results (red circles) based on the model. The dashed lines in both panels have a slope -2.08 as the analytic results in Eqs. (17) and (31) suggested. The simulated network is generated by the model with the parameters $N = 100000$, $m \approx 8.07$ and $p \approx 0.39$ as determined by the mean degree $\langle k \rangle$ and reciprocity of the Flickr social network. Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

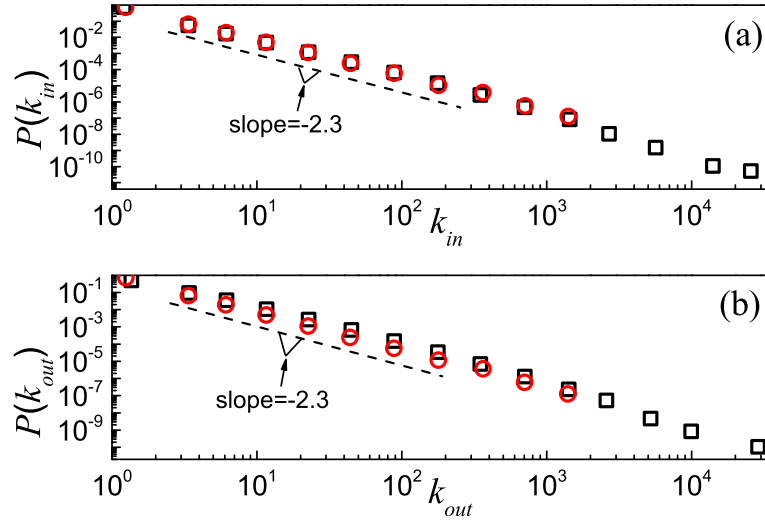


Figure S3. Indegree (a) and outdegree (b) distributions of the YouTube social network (black squares) and simulation results (red circles) based on the model. The dashed lines in both panels have a slope -2.3 as the analytic results in Eqs. (17) and (31) suggested. The simulated network is generated by the model with the parameters $N = 100000$, $m \approx 4.34$ and $p \approx 0.08$ as determined by the mean degree $\langle k \rangle$ and reciprocity of the YouTube social network. Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

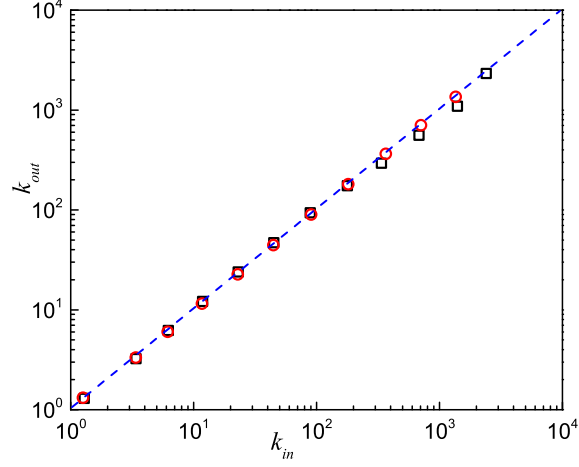


Figure S4. Relationship between the indegree and the outdegree of nodes in the Slashdot social network and the model. Results of the Slashdot social network (black squares) and simulation results (red circles) based on the model are shown. The blue dash line represents the relation function $k_{in} = k_{out}$. Data points are averages over the logarithmic bins of the indegree k_{in} .

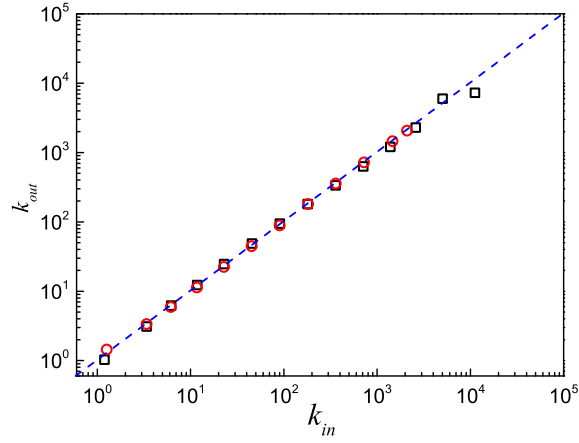


Figure S5. Relationship between the indegree and the outdegree of nodes in the Flickr social network and the model. Results of the Flickr social network (black squares) and simulation results (red circles) based on the model are shown. The blue dash line represents the relation function $k_{in} = k_{out}$. Data points are averages over the logarithmic bins of the indegree k_{in} .

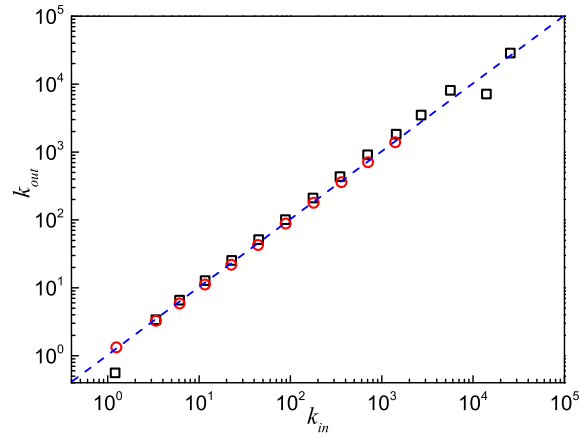


Figure S6. Relationship between the indegree and the outdegree of nodes in the YouTube social network and the model. Results of the YouTube social network (black squares) and simulation results (red circles) based on the model are shown. The blue dash line represents the relation function $k_{in} = k_{out}$. Data points are averages over the logarithmic bins of the indegree k_{in} .

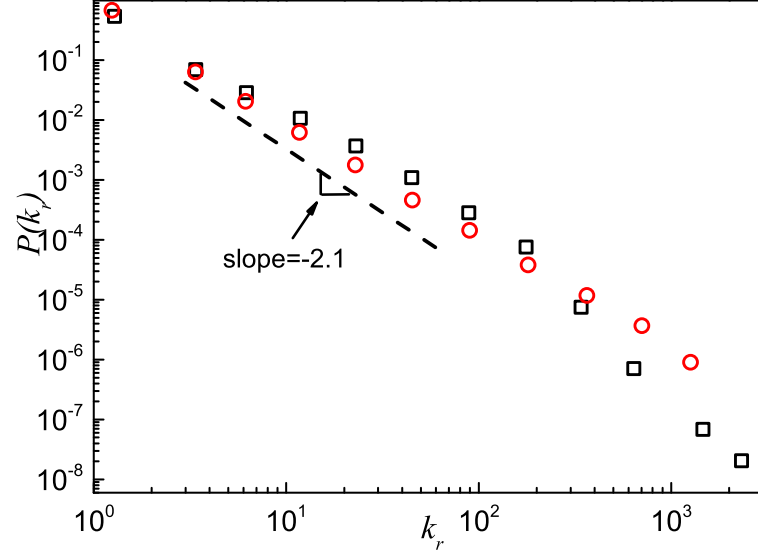


Figure S7. Reciprocal degree distributions of the Slashdot social network and the model. Results of the Slashdot social network (black squares) and simulation results (red circles) based on the model are shown. Analytic treatment (see Eqs. (17) and (31)) suggests a scaling behavior with an exponent -2.1 , as shown by the dash line. Data points are averages over the logarithmic bins of the reciprocal degree k_r .

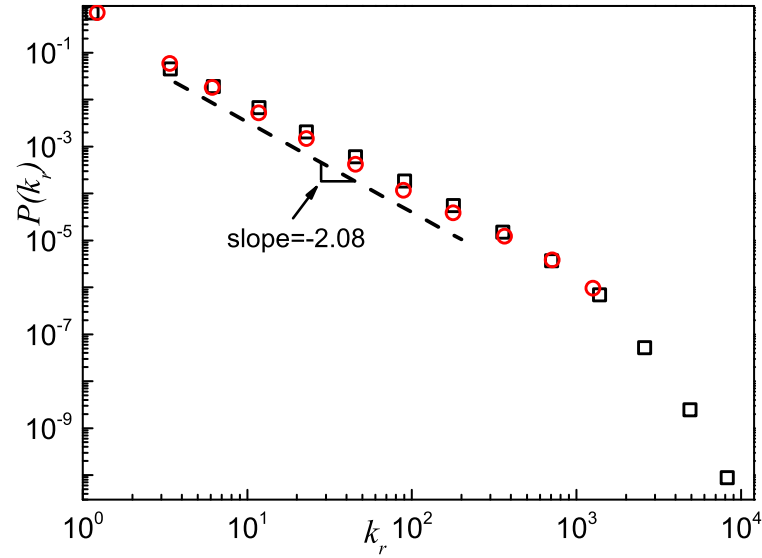


Figure S8. Reciprocal degree distributions of the Flickr social network and the model. Results of the Flickr social network (black squares) and simulation results (red circles) based on the model are shown. Analytic treatment (see Eqs. (17) and (31)) suggests a scaling behavior with an exponent -2.08 , as shown by the dash line. Data points are averages over the logarithmic bins of the reciprocal degree k_r .

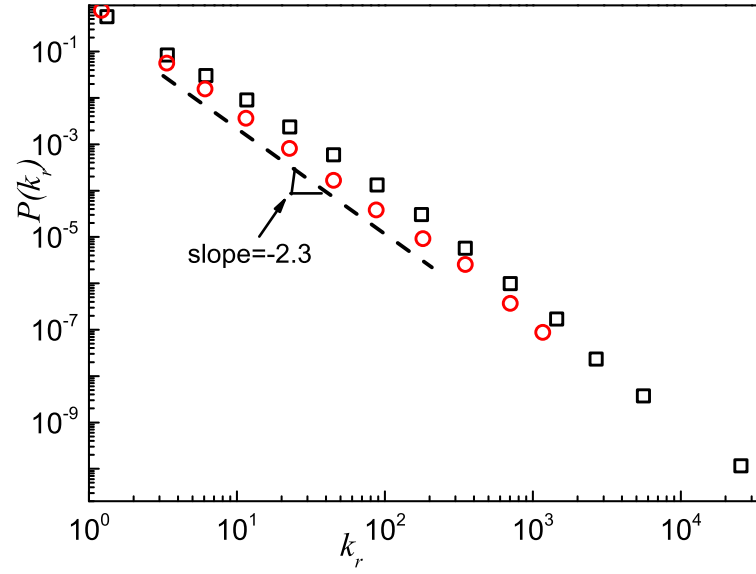


Figure S9. Reciprocal degree distributions of the YouTube social network and the model. Results of the YouTube social network (black squares) and simulation results (red circles) based on the model are shown. Analytic treatment (see Eqs. (17) and (31)) suggests a scaling behavior with an exponent -2.3 , as shown by the dash line. Data points are averages over the logarithmic bins of the reciprocal degree k_r .

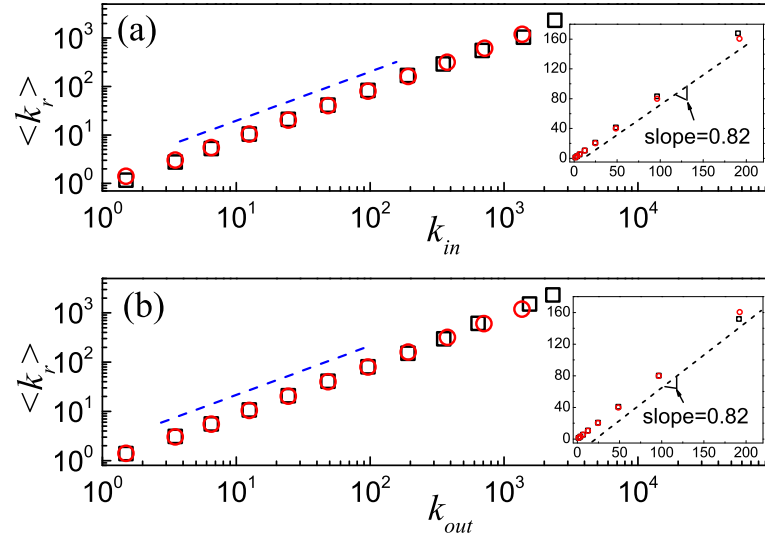


Figure S10. Mean reciprocal degree of nodes with (a) the same indegree and (b) the same outdegree in the Slashdot social network and in the model. Results of the Slashdot social network (black squares) and simulation results (red circles) based on the model are shown in a log-log scale in the main panels. Analytic treatment suggests that $\langle k_r \rangle$ is linearly dependent on k_{in} and k_{out} , and the blue dash lines of slope 1 show its dependence. The inset in each panel shows the results in a linear scale and the dash line has a slope of 0.82, as given by Eqs. (20) and (31). Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

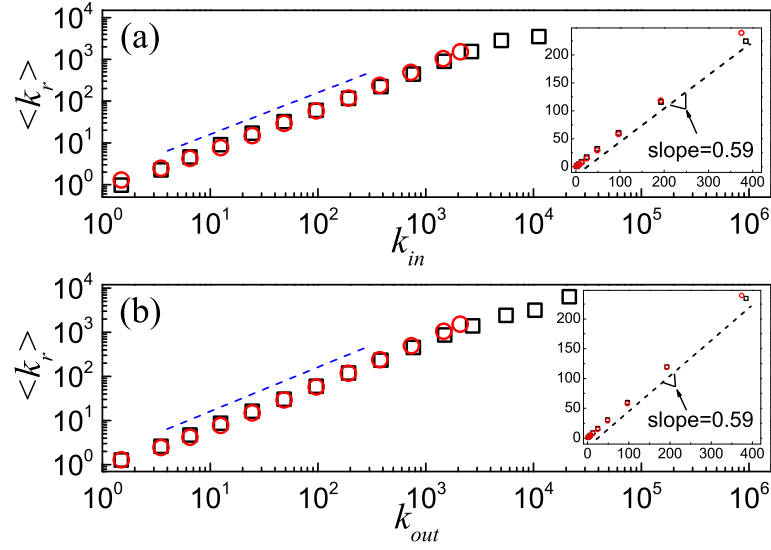


Figure S11. Mean reciprocal degree of nodes with (a) the same indegree and (b) the same outdegree in the Flickr social network and in the model. Results of the Flickr social network (black squares) and simulation results (red circles) based on the model are shown in a log-log scale in the main panels. Analytic treatment suggests that $\langle k_r \rangle$ is linearly dependent on k_{in} and k_{out} , and the blue dash lines of slope 1 show its dependence. The inset in each panel shows the results in a linear scale and the dash line has a slope of 0.59, as given by Eqs. (20) and (31). Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

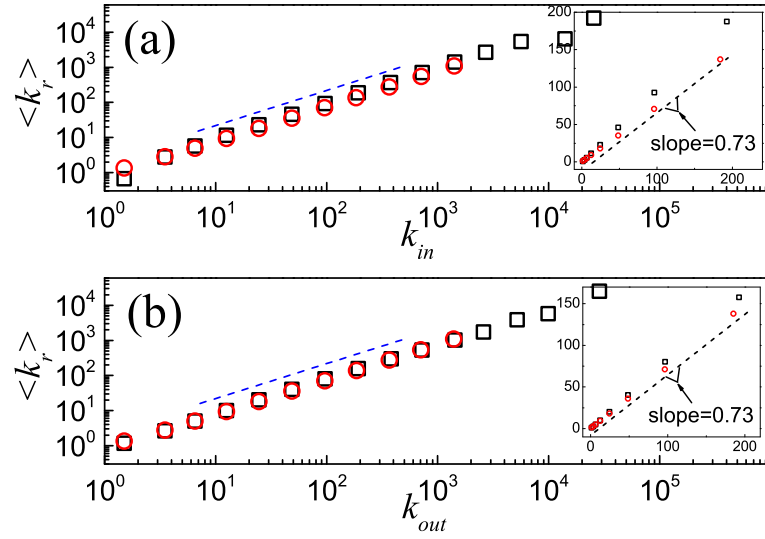


Figure S12. Mean reciprocal degree of nodes with (a) the same indegree and (b) the same outdegree in the YouTube social network and in the model. Results of the YouTube social network (black squares) and simulation results (red circles) based on the model are shown in a log-log scale in the main panels. Analytic treatment suggests that $\langle k_r \rangle$ is linearly dependent on k_{in} and k_{out} , and the blue dash lines of slope 1 show its dependence. The inset in each panel shows the results in a linear scale and the dash line has a slope of 0.73, as given by Eqs. (20) and (31). Data points are averages over the logarithmic bins of the indegree k_{in} and outdegree k_{out} , respectively.

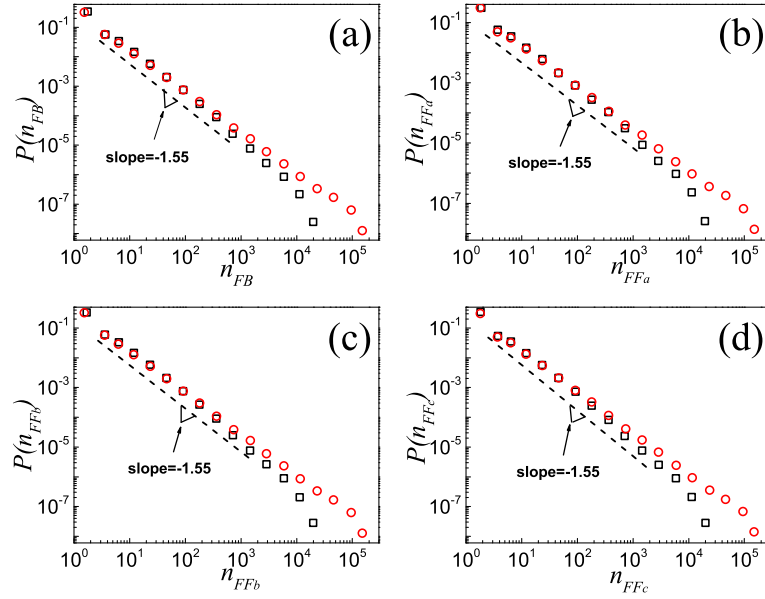


Figure S13. Distributions of four basic closed triples in the slashdot social network and the model. Distributions of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops in the Slashdot social network (black squares) and in the simulated network based on the model (red circles). Analytic treatment (see Eqs. (30) and (31)) suggests a scaling behavior with an exponent -1.55 , as shown by the dash lines. Data points are averages over the logarithmic bins of the n_{FB} , n_{FFa} , n_{FFb} and n_{FFc} , respectively.

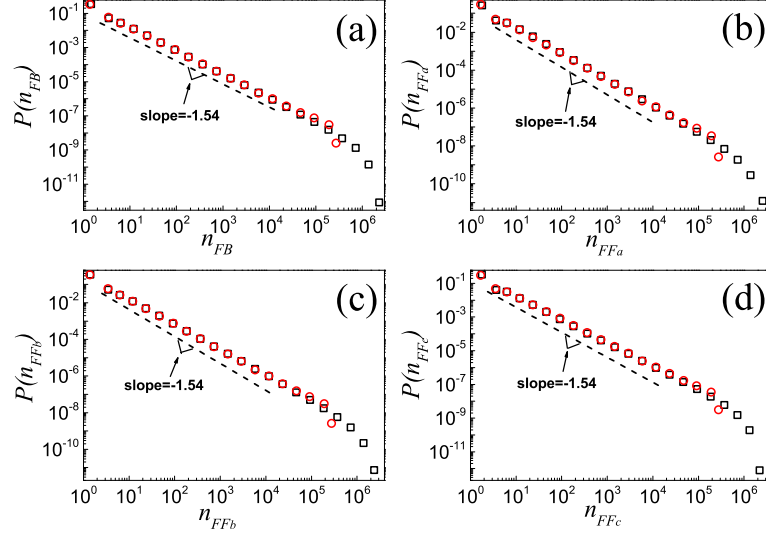


Figure S14. Distributions of four basic closed triples in the Flickr social network and the model. Distributions of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops in the Flickr social network (black squares) and in the simulated network based on the model (red circles). Analytic treatment (see Eqs. (30) and (31)) suggests a scaling behavior with an exponent -1.54 , as shown by the dash lines. Data points are averages over the logarithmic bins of the n_{FB} , n_{FF_a} , n_{FF_b} and n_{FF_c} , respectively.

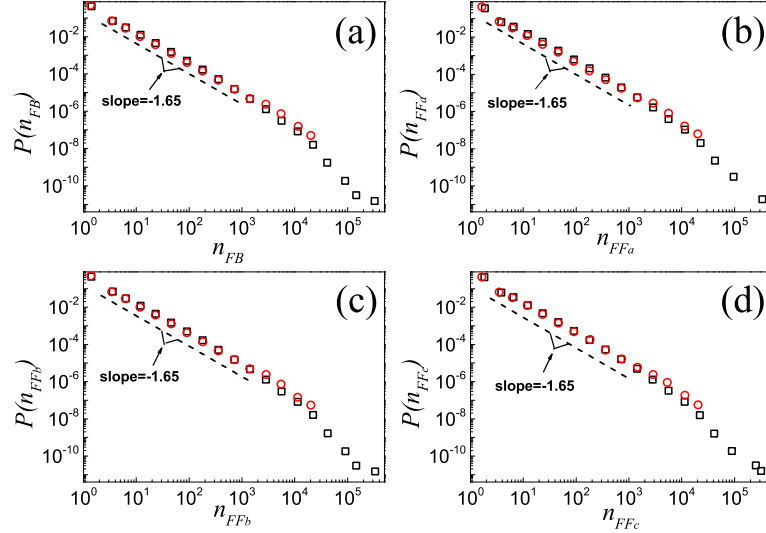


Figure S15. Distributions of four basic closed triples in the YouTube social network and the model. Distributions of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops in the YouTube social network (black squares) and in the simulated network based on the model (red circles). Analytic treatment (see Eqs. (30) and (31)) suggests a scaling behavior with an exponent -1.65 , as shown by the dash lines. Data points are averages over the logarithmic bins of the n_{FB} , n_{FF_a} , n_{FF_b} and n_{FF_c} , respectively.

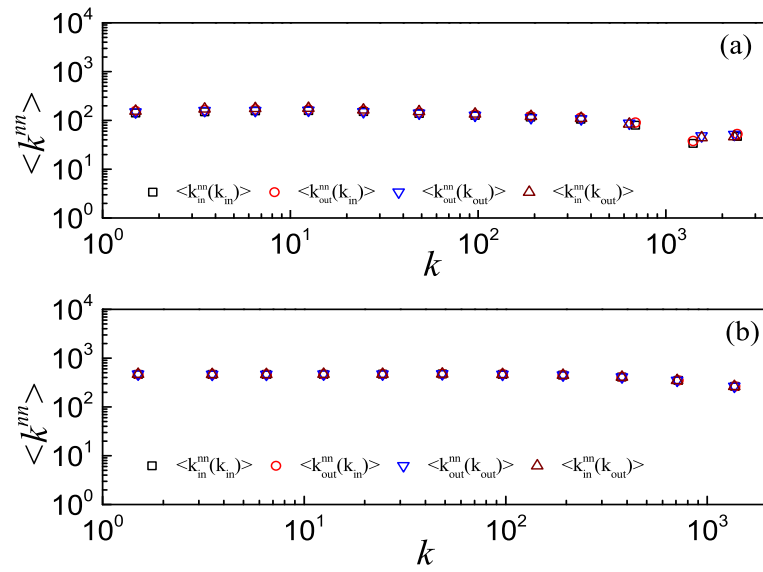


Figure S16. Degree correlations in the Slashdot social network and the model. Results of degree correlations as measured by four quantities corresponding to the average nearest neighbor degree $\langle k^{nn}(k_{in}) \rangle$ (squares), $\langle k^{nn}(k_{in}) \rangle$ (circles), $\langle k^{nn}(k_{out}) \rangle$ (triangles), and $\langle k^{nn}(k_{out}) \rangle$ (inverted triangles) for (a) Slashdot social network and (b) simulated network based on the model. Data points are averages over the logarithmic bins of the indegree k_{in} or outdegree k_{out} .

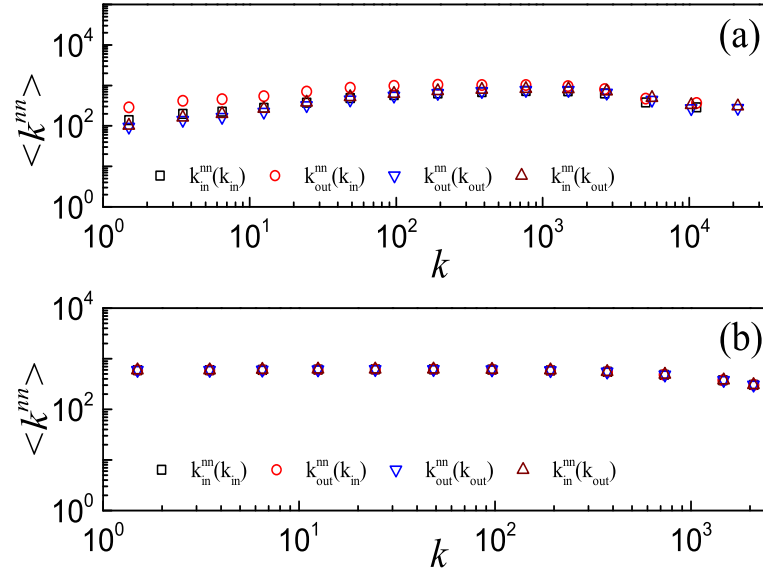


Figure S17. Degree correlations in the Flickr social network and the model. Results of degree correlations as measured by four quantities corresponding to the average nearest neighbor degree $\langle k^{nn}_{in}(k_{in}) \rangle$ (squares), $\langle k^{nn}_{out}(k_{in}) \rangle$ (circles), $\langle k^{nn}_{out}(k_{out}) \rangle$ (triangles), and $\langle k^{nn}_{in}(k_{out}) \rangle$ (inverted triangles) for (a) Flickr social network and (b) simulated network based on the model. Data points are averages over the logarithmic bins of the indegree k_{in} or outdegree k_{out} .

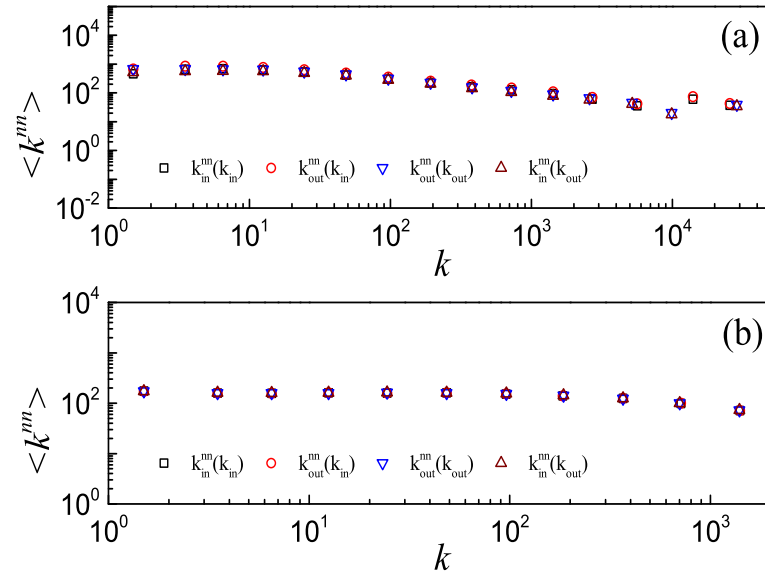


Figure S18. Degree correlations in the YouTube social network and the model. Results of degree correlations as measured by four quantities corresponding to the average nearest neighbor degree $\langle k^{nn}_{in}(k_{in}) \rangle$ (squares), $\langle k^{nn}_{out}(k_{in}) \rangle$ (circles), $\langle k^{nn}_{out}(k_{out}) \rangle$ (triangles), and $\langle k^{nn}_{in}(k_{out}) \rangle$ (inverted triangles) for (a) YouTube social network and (b) simulated network based on the model. Data points are averages over the logarithmic bins of the indegree k_{in} or outdegree k_{out} .

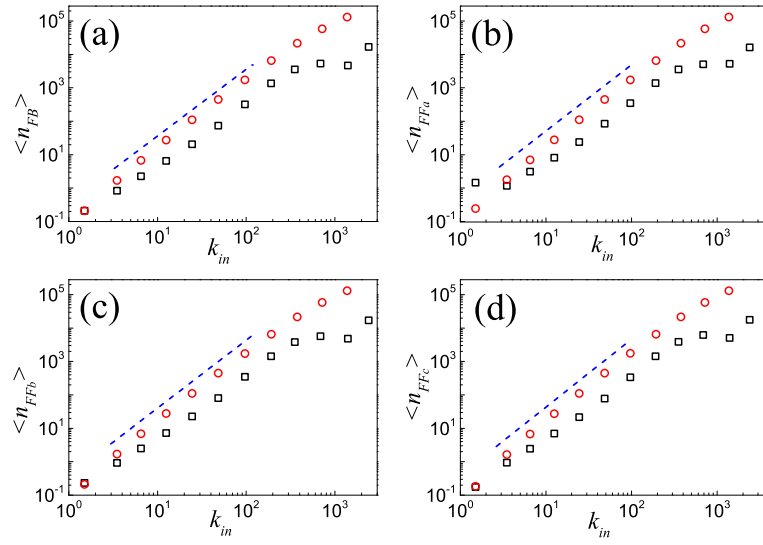


Figure S19. Mean number of the four closed triples for nodes with the same indegree in the Slashdot social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same indegree are shown for the Slashdot social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the indegree k_{in} .

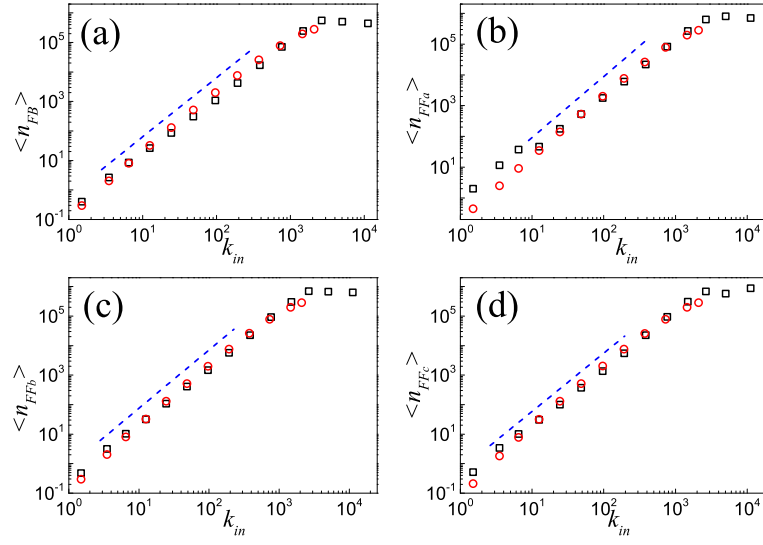


Figure S20. Mean number of the four closed triples for nodes with the same indegree in the Flickr social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same indegree are shown for the Flickr social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the indegree k_{in} .

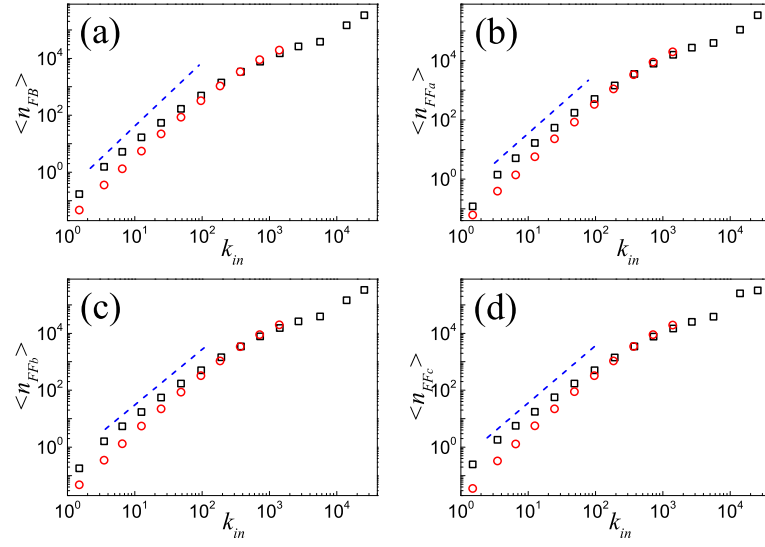


Figure S21. Mean number of the four closed triples for nodes with the same indegree in the YouTube social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same indegree are shown for the YouTube social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the indegree k_{in} .

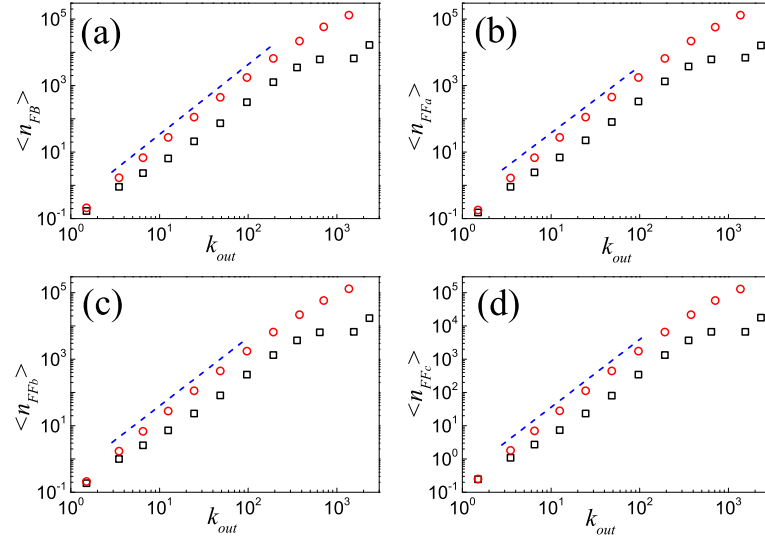


Figure S22. Mean number of the four closed triples for nodes with the same outdegree in the Slashdot social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same outdegree are shown for the Slashdot social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the outdegree k_{out} .

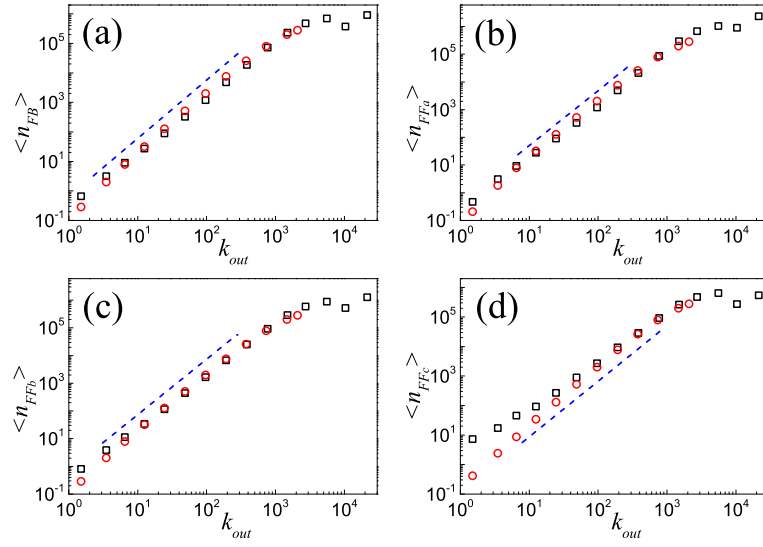


Figure S23. Mean number of the four closed triples for nodes with the same outdegree in the Flickr social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same outdegree are shown for the Flickr social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the outdegree k_{out} .

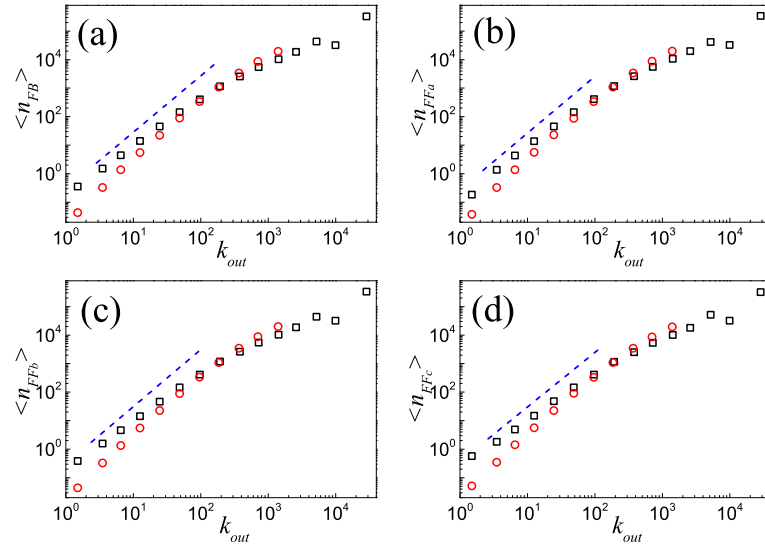


Figure S24. Mean number of the four closed triples for nodes with the same outdegree in the YouTube social network and the model. Results for the mean number of closed triples corresponding to (a) FB , (b) FF_a , (c) FF_b , and (d) FF_c loops for nodes with the same outdegree are shown for the YouTube social network (black squares) and simulated network (red circles) based on the model. Analytic treatment (see Eq. (28)) gives a scaling behavior with an exponent 2, as indicated by the dash line. Data points are averages over the logarithmic bins of the outdegree k_{out} .